

GS.TS.NGUYỄN BÙI XUÂN PHONG

THỐNG KÊ VÀ ỨNG DỤNG

TRONG BƯỞI CHÍNH, VIỆN THÔNG

UYÊN
LIỆU



NHÀ XUẤT BẢN BƯU ĐIỆN

THỐNG KÊ VÀ ỨNG DỤNG

TRONG BƯU CHÍNH, VIỄN THÔNG

GS.TS. NGUYỄN BÙI XUÂN PHONG

THỐNG KÊ VÀ ỨNG DỤNG TRONG BƯU CHÍNH, VIỄN THÔNG

NHÀ XUẤT BẢN BƯU ĐIỆN

Hà Nội, tháng 9 - 2005

Mã số: KK 06 HM 05

LỜI NHÀ XUẤT BẢN

Thống kê, đặc biệt là thống kê ứng dụng là một môn học không thể thiếu trong chương trình đào tạo ngành Kinh tế và Quản trị kinh doanh Bưu chính, Viễn thông. Thống kê là hoạt động quan trọng, là công cụ không thể thiếu trong quản trị doanh nghiệp và trong quản lý kinh tế.

Với mục đích cung cấp cho bạn đọc một phương pháp tổng hợp, phân tích số liệu một cách khoa học và hiệu quả cao phục vụ cho những doanh nghiệp hoạt động sản xuất kinh doanh trong lĩnh vực Bưu chính, Viễn thông; Nhà xuất bản Bưu điện xuất bản cuốn sách ***“Thống kê và ứng dụng trong Bưu chính, Viễn thông”*** do GS.TS. NGUYỄN Bùi Xuân Phong - Phó Trưởng khoa Quản trị Kinh doanh, Học viện Công nghệ Bưu chính Viễn thông biên soạn. Nội dung cuốn sách gồm 07 chương đề cập những kiến thức rất thiết thực về thống kê và những ứng dụng trong Bưu chính, Viễn thông như: Khái niệm về thống kê; Phương pháp trình bày số liệu thống kê; Điều tra số liệu chọn mẫu; Tương quan và hồi quy; Chỉ số trong thống kê...

Cuốn sách là tư liệu tham khảo cần thiết cho các lãnh đạo, các cán bộ hoạch định chính sách, cán bộ quản lý của các cơ quan, đơn vị thuộc Bộ Bưu chính, Viễn thông, Tổng Công ty Bưu chính Viễn thông Việt Nam, các Sở Bưu chính Viễn thông, đặc biệt là các doanh nghiệp kinh doanh dịch vụ Bưu chính, Viễn

thông và Công nghệ thông tin; đồng thời cuốn sách là tài liệu tham khảo hữu ích đối với các giáo viên, sinh viên Học viện Công nghệ Bưu chính Viễn thông, các cơ sở đào tạo chuyên ngành Bưu điện và bạn đọc quan tâm đến vấn đề này.

Các ý kiến góp ý của quý vị bạn đọc, xin gửi về địa chỉ: ***Nhà xuất bản Bưu điện***, 18 Nguyễn Du, Hà Nội; Điện thoại: ***04. 9430202***; Fax: ***04. 9431285***.

Trân trọng cảm ơn./.

Hà Nội, tháng 9 năm 2005

NHÀ XUẤT BẢN BƯU ĐIỆN

Chương 1

MỘT SỐ KHÁI NIỆM VỀ THỐNG KÊ

1.1. THỐNG KÊ HỌC LÀ GÌ?

Thuật ngữ thống kê có hai nghĩa:

- *Theo nghĩa thứ nhất*, thống kê là các con số được ghi chép để phản ánh các hiện tượng tự nhiên, kỹ thuật, kinh tế và xã hội. Chẳng hạn sản lượng dịch vụ, doanh thu kinh doanh bưu chính viễn thông (BCVT)...

- *Theo nghĩa thứ hai*, thống kê được hiểu là hệ thống các phương pháp ghi chép, thu thập và phân tích các con số về những hiện tượng tự nhiên, kỹ thuật, kinh tế và xã hội để tìm hiểu bản chất và tìm quy luật vốn có của những hiện tượng ấy. Chẳng hạn như làm thế nào để có được các con số về lao động của một doanh nghiệp BCVT ở một thời điểm nào đó và nghiên cứu sâu vào cơ cấu lứa tuổi, giới tính, trình độ, nghề nghiệp... từ đó mà có cách đánh giá đúng đắn về thực trạng lao động, giúp cho việc sử dụng có hiệu quả lao động, đồng thời có chính sách đào tạo, tuyển dụng lao động nhằm hoàn thành nhiệm vụ chính trị của doanh nghiệp.

Thống kê học chính là *khoa học nghiên cứu hệ thống các phương pháp thu thập, xử lý và phân tích các con số (mặt lượng) của những hiện tượng số lớn để tìm hiểu bản chất và tính quy luật vốn có của chúng (mặt chất) trong những điều kiện, địa điểm và thời gian cụ thể.*

Hiện tượng bao giờ cũng có hai mặt chất và lượng không tách rời nhau. Chất của hiện tượng giúp ta phân biệt hiện tượng này với hiện tượng khác, đồng thời bộc lộ những khía cạnh sâu kín của hiện tượng. Nhưng chất không tồn tại độc lập mà được biểu hiện qua lượng, với những cách xử lý mặt lượng đó một cách khoa học. Sở dĩ cần phải xử lý mặt lượng mới tìm hiểu được mặt chất là vì mặt chất của hiện tượng thường bị che khuất dưới các tác động ngẫu nhiên. Phải thông qua tổng hợp mặt lượng của số lớn đơn vị cấu thành hiện tượng, tác động của các yếu tố ngẫu nhiên mới được bù trừ và triệt tiêu. Hơn nữa, cũng còn phải sử dụng các phương pháp phân tích số liệu thích hợp, bản chất của hiện tượng mới dần dần bộc lộ qua tính quy luật thống kê. Về thực chất, tính quy luật thống kê là sự biểu hiện về lượng của các quy luật phát sinh, phát triển của hiện tượng. Tính quy luật này không có tính chất chung chung mà rất cụ thể theo các điều kiện, địa điểm và thời gian cụ thể. Đó chính là đặc trưng của thống kê học, làm cho nó khác với toán học.

Tính quy luật thống kê có ý nghĩa rất quan trọng đối hoạt động kinh doanh, trong đó có hoạt động kinh doanh BCVT, vì nó cho biết mối liên hệ giữa các hiện tượng, xu thế phát triển của hiện tượng cũng như các dao động chu kỳ của hiện tượng đó, quy luật phân phối của các tổng thể chứa đựng hiện tượng đang nghiên cứu.

1.2. TỔNG THỂ THỐNG KÊ VÀ ĐƠN VỊ TỔNG THỂ THỐNG KÊ

Thống kê nghiên cứu mặt lượng của hiện tượng kinh tế xã hội số lớn phải xác định phạm vi hiện tượng được nghiên cứu cụ thể. Để chỉ đối tượng nghiên cứu cụ thể, người ta dùng khái niệm

tổng thể. Tổng thể thống kê là hiện tượng kinh tế xã hội số lớn, gồm những đơn vị (hoặc phân tử, hiện tượng) cá biệt cần được quan sát, phân tích mặt lượng của chúng để tìm hiểu bản chất và tính quy luật vốn có của chúng (mặt chất) trong những điều kiện, địa điểm và thời gian cụ thể. Chẳng hạn toàn bộ cán bộ, công nhân viên của một bưu cục vào một thời gian nào đó là một tổng thể. Các bưu cục thuộc mạng bưu chính vào một thời gian xác định cũng là một tổng thể...

Có trường hợp các đơn vị cấu thành tổng thể, có thể thấy được bằng trực quan. Tổng thể bao gồm các đơn vị như vậy được gọi là tổng thể bộc lộ như các tổng thể nêu trên. Tổng thể các đơn vị cấu thành nó, không thể nhận biết được bằng trực quan là tổng thể tiềm ẩn như tổng thể những cán bộ công nhân viên ưa chuộng nghệ thuật sân khấu, tổng thể những người mê tín dị đoan, tổng thể những người trung thành với Tổ quốc...

Các đơn vị tổng thể có thể giống nhau trên một số đặc điểm, các đặc điểm còn lại khác nhau. Do đó, tùy theo mục đích nghiên cứu mà phân biệt tổng thể đồng chất hay không đồng chất. Tổng thể đồng chất bao gồm các đơn vị giống nhau về một số đặc điểm chủ yếu có liên quan tới mục đích nghiên cứu, tổng thể không đồng chất bao gồm các đơn vị khác nhau về các đặc điểm, các loại hình.

Tổng thể bao gồm tất cả các đơn vị thuộc phạm vi nghiên cứu tổng thể chung, chỉ bao gồm một bộ phận đơn vị trong đó là tổng thể bộ phận.

Xác định tổng thể để đáp ứng mục đích nghiên cứu thống kê phải trên cơ sở phân tích lý luận kinh tế, chính trị hoặc xã hội,

định nghĩa rõ tổng thể. Định nghĩa tổng thể không những phải giới hạn về thực thể (tổng thể là gì) mà còn phải giới hạn về thời gian và không gian (tổng thể tồn tại vào thời gian nào, ở đâu). Xác định tổng thể chính xác không dễ dàng vì có những hiện tượng có thể tương tự về hình thức, nhưng lại khác hẳn về nội dung. Chính vì vậy phải phân tích lý luận để thấy rõ nội dung của hiện tượng. Xác định tổng thể thống kê không chính xác sẽ lãng phí sức người và tiền của trong nghiên cứu, không đủ cơ sở để hiểu đúng bản chất cụ thể của hiện tượng.

Định nghĩa tổng thể làm rõ đặc trưng cơ bản chung của hiện tượng kinh tế xã hội, số lớn phù hợp với mục đích nghiên cứu. Thông qua việc phân tích lý luận và thực tế phải làm rõ tổng thể gồm những hiện tượng (phần tử) cá biệt nào. Hiện tượng cá biệt này là đơn vị tổng thể. Tất cả các đơn vị tổng thể chỉ giống nhau trên một số mặt, còn các mặt khác không giống nhau. Cho nên trong thực tế phải nêu rõ ràng những hiện tượng cá biệt nào được kể là đơn vị tổng thể. Trong những trường hợp khó khăn cho việc giới hạn, người ta phải lập một danh mục các đơn vị hoặc trong giải thích cần xác định rõ phạm vi nào của các đơn vị thuộc tổng thể.

Đơn vị tổng thể bao giờ cũng có đơn vị tính toán phù hợp. Xác định đơn vị tổng thể là việc cụ thể hóa tổng thể. Đơn vị tổng thể là xuất phát điểm của quá trình nghiên cứu thống kê vì nó có mặt lượng mà ta cần nghiên cứu, cho nên xác định đơn vị tổng thể cũng quan trọng như xác định tổng thể.

1.3. TIÊU THỨC THỐNG KÊ

Nghiên cứu thống kê phải dựa vào các đặc điểm của đơn vị tổng thể. Đơn vị tổng thể có nhiều đặc điểm, tùy theo mục đích

nghiên cứu, một số đặc điểm của đơn vị tổng thể được chọn ra để nghiên cứu. Các đặc điểm này được gọi là *các tiêu thức*. Ví dụ mỗi cán bộ công nhân viên của bưu cục có các tiêu thức: tên, tuổi, giới tính, trình độ văn hóa, nghề nghiệp, nơi ở... Mỗi bưu cục trong tổng thể có tiêu thức: tên bưu cục, địa chỉ, số lượng cán bộ công nhân viên...

Đơn vị tổng thể được làm rõ đặc trưng của nó qua các tiêu thức: thực thể, thời gian và không gian.

1.3.1. Tiêu thức thực thể

Tiêu thức thực thể nêu lên bản chất của đơn vị tổng thể và nó biến đổi trong bản chất này. Các tiêu thức: giới tính, tuổi, trình độ văn hóa, nghề nghiệp, số lượng cán bộ công nhân viên chức là các tiêu thức thực thể. Theo nội dung của nó, tiêu thức thực thể gồm hai loại: thuộc tính và số lượng.

Tiêu thức thuộc tính không có biểu hiện trực tiếp là các con số, như tiêu thức giới tính, trình độ văn hóa, nghề nghiệp... Tiêu thức thuộc tính có biểu hiện trực tiếp và gián tiếp, như giới tính có biểu hiện trực tiếp là nam và nữ. Tiêu thức đời sống vật chất có biểu hiện gián tiếp là lương tiêu dùng lương thực, thực phẩm theo đầu người, diện tích nhà ở theo đầu người. Các biểu hiện gián tiếp của tiêu thức thuộc tính còn được gọi là các chỉ báo thống kê.

Tiêu thức thuộc tính không có biểu hiện trực tiếp là con số, nên còn được gọi là tiêu thức phi lượng hóa.

Tiêu thức số lượng có biểu hiện trực tiếp là con số (gọi là lượng biến). Nó là kết quả của quá trình quan sát (cân đo, đong đếm) như tuổi đời, tuổi nghề, số lượng điện thoại, số bưu cục...

Tiêu thức số lượng còn gọi là tiêu thức lượng hóa vì nó có biểu hiện trực tiếp là con số.

Tiêu thức thực thể khi chỉ có hai biểu hiện không trùng nhau trên một đơn vị tổng thể, được gọi là *tiêu thức thay phiên*, như giới tính (nam và nữ), chất lượng sản phẩm dịch vụ (tốt và xấu). Tiêu thức thực thể có ba loại biểu hiện trở lên có thể trở thành *tiêu thức thay phiên*. Ví dụ như số lượng cán bộ công nhân viên của bưu cục nêu trên có nhiều biểu hiện nhưng rút gọn thành hai biểu hiện, hoặc các biểu hiện của tiêu thức trình độ văn hóa có thể rút gọn thành hai biểu hiện: chưa tốt nghiệp phổ thông trung học và tốt nghiệp phổ thông trung học trở lên. Những trường hợp này được tiến hành khi người ta chỉ quan tâm đến một biểu hiện nào đó xuất hiện hay không xuất hiện trên đơn vị tổng thể.

Tiêu thức thực thể có phù hợp nhiều hay ít với việc đáp ứng mục đích nghiên cứu là tùy thuộc vào việc chọn những tiêu thức nào cho nghiên cứu.

1.3.2. Tiêu thức thời gian

Tiêu thức thời gian nêu hiện tượng kinh tế xã hội theo sự xuất hiện của nó vào thời gian nào. Những biểu hiện của tiêu thức thời gian là phút, giờ, ngày, tháng, năm. Thời hạn có giá trị của các chỉ dẫn về đối tượng nghiên cứu và những đơn vị tổng thể, về sự phân phối chúng trong một thời gian cũng như về sự thay đổi từ thời kỳ này tới thời kỳ khác được khẳng định qua tiêu thức thời gian. Ví dụ tổng số máy điện thoại có đến ngày 31/12/2001 là 4.301.120, trong đó 70,01% máy cố định; 29,99% máy di động.

1.3.3. Tiêu thức không gian

Tiêu thức không gian nêu phạm vi lãnh thổ bao trùm của đối tượng nghiên cứu và sự xuất hiện theo địa điểm của các đơn vị tổng thể. Những biểu hiện của nó được chỉ ra nhờ sự phân định về mặt quản lý hành chính hoặc theo điều kiện tự nhiên, phân vùng kinh tế... Nghiên cứu thống kê theo tiêu thức không gian có ý nghĩa quan trọng trước hết là gắn với tiêu thức thực thể để quan sát sự phân phối về mặt lãnh thổ của các đơn vị tổng thể.

Các tiêu thức này góp phần vào việc khẳng định đơn vị tổng thể cũng như tổng thể vì chúng nêu rõ các mặt và tính chất nhất định của đơn vị tổng thể cũng như của tổng thể; nhờ đó có thể phân biệt đơn vị này với đơn vị khác cũng như tổng thể này với tổng thể kia.

1.4. CHỈ TIÊU THỐNG KÊ

Nghiên cứu thống kê không chỉ phản ánh lượng và chất của hiện tượng kinh tế xã hội cá biệt mà còn phản ánh mặt lượng và chất của hiện tượng kinh tế xã hội số lớn trong điều kiện thời gian và địa điểm cụ thể. Tính chất của các hiện tượng cá biệt được khái quát hóa trong chỉ tiêu thống kê. Do đó chỉ tiêu thống kê chỉ ra những mối quan hệ cần thiết, cái chung của tất cả các đơn vị hoặc của nhóm đơn vị. Ngoài ra chỉ tiêu thống kê còn phản ánh các mối quan hệ tồn tại khách quan, nhưng cũng không tự bộc lộ ra để hiểu trực tiếp là mối quan hệ. Phải điều tra mặt lượng của đơn vị cá biệt và từ đó phát hiện ý nghĩa theo số lượng của mối quan hệ bằng chỉ tiêu thống kê. Các mối quan hệ này sẽ được nghiên cứu trong các phương pháp thuộc các chương sau.

Chỉ tiêu thống kê có hai mặt: khái niệm và con số. Khái niệm của chỉ tiêu thống kê bao gồm định nghĩa và giới hạn về thực thể, thời gian và không gian của hiện tượng kinh tế xã hội. Mặt này chỉ rõ nội dung của chỉ tiêu thống kê. Con số của chỉ tiêu thống kê là trị số được phát hiện với đơn vị tính toán phù hợp và nó nêu lên mức độ của chỉ tiêu thống kê. Ví dụ doanh thu phát sinh năm 2001 là 18.617,15 tỷ đồng ; trong đó doanh thu phát sinh 2001 là mặt khái niệm của chỉ tiêu, 18.617,15 tỷ đồng là mặt con số của chỉ tiêu.

Theo nội dung, chỉ tiêu thống kê biểu hiện quy mô, cơ cấu, sự phát triển và mối quan hệ của hiện tượng số lớn trong điều kiện thời gian và địa điểm cụ thể.

Căn cứ vào nội dung có thể chia các chỉ tiêu thống kê thành hai loại: khối lượng và chất lượng. Chỉ tiêu thống kê khối lượng biểu hiện quy mô của tổng thể như số cán bộ công nhân viên, số máy điện thoại, khối lượng sản phẩm dịch vụ. Chỉ tiêu thống kê chất lượng biểu hiện trình độ phổ biến, mối quan hệ của tổng thể cũng như giá thành sản phẩm dịch vụ. Việc phân loại này nhằm đáp ứng yêu cầu của một số phương pháp phân tích thống kê.

1.5. CÁC LOẠI THANG ĐO

Theo tính chất của việc đo lường, người ta thường có 04 loại thang đo:

- **Thang đo định danh** (hay là đặt tên) là đánh số các biểu hiện cùng loại của một tiêu thức. Như giới tính biểu hiện “nam” được đánh số 1 và “nữ” đánh số 2. Giữa các con số ở đây không có quan hệ hơn, kém, cho nên các phép tính với chúng đều vô nghĩa. Loại thang đo này dùng để đếm tần số của biểu hiện tiêu thức.

- **Thang đo thứ bậc** cũng là thang đo định danh, nhưng giữa các biểu hiện tiêu thức có quan hệ thứ bậc, hơn kém. Sự chênh lệch giữa các biểu hiện không nhất thiết phải bằng nhau, như Huân chương có ba hạng: Nhất, Nhì và Ba. Hạng Nhất hơn hạng Nhì, hạng Nhì hơn hạng Ba. Trình độ văn hóa phổ thông có ba cấp: Tiểu học, Trung học cơ sở và Trung học phổ thông. Trung học phổ thông hơn Trung học cơ sở, Trung học cơ sở hơn Tiểu học. Con số có trị số lớn hơn không có nghĩa ở bậc cao hơn và ngược lại, mà do sự quy định. Thang đo loại này được sử dụng để tính toán đặc trưng chung của tổng thể một cách tương đối, trong một số trường hợp như tính cấp bậc bình quân của một doanh nghiệp, một đơn vị, bộ phận.

- **Thang đo khoảng** là thang đo thứ bậc có các khoảng cách đều nhau và có thể đánh giá sự khác biệt giữa các biểu hiện bằng thang đo loại này. Việc cộng trừ các con số có ý nghĩa, có thể tính các đặc trưng chung như số bình quân, phương sai. Yêu cầu có khoảng cách đều là đặt ra đối với thang đo, còn đối với biểu hiện của tiêu thức được đo không nhất thiết phải bằng nhau.

- **Thang đo tỷ lệ** là thang đo khoảng với một điểm không (0) tuyệt đối (điểm gốc) để có thể so sánh được tỷ lệ giữa các trị số đo. Với thang đo loại này, có thể đo lường các biểu hiện của tiêu thức như các đơn vị đo lường vật lý thông thường (kg, m,...) và thực hiện được tất cả các phép tính với trị số đo.

Chương 2

CÁC PHƯƠNG PHÁP TRÌNH BÀY SỐ LIỆU THỐNG KÊ

2.1. TRÌNH BÀY SỐ LIỆU BẰNG BẢNG THỐNG KÊ

2.1.1. Đối với tiêu thức thuộc tính

Khi trình bày số liệu bằng bảng thống kê đối với tiêu thức thuộc tính, có thể gặp những trường hợp sau:

Trường hợp 1: Chỉ xét những biểu hiện của một tiêu thức.

Trong trường hợp này bảng thống kê có những dạng sau:

Bảng 2.1

Tiêu thức A_i	Tần số f_i	Tần suất d_i
A_1	f_1	d_1
A_2	f_2	d_2
...	...	...
A_k	f_k	d_k
Cộng	Σf_i	$\Sigma d_i = 1$

Cột "Tiêu thức A_i " ứng với mỗi dòng ta ghi các biểu hiện của tiêu thức A ($i = 1, 2, \dots k$).

Cột "Tần số f_i " ghi số lần quan sát được ứng với từng biểu hiện A_i .

Cột "Tần suất d_i " ghi tần suất tương ứng là tần số tính bằng số tương đối, được xác định theo công thức:

$$d_i = \frac{f_i}{\sum f_i}$$

Ví dụ: Doanh thu của một số Bưu điện tỉnh như sau: Bưu điện tỉnh A là 4.000 triệu đồng; Bưu điện tỉnh B: 2.400; Bưu điện tỉnh C: 11.000 và Bưu điện tỉnh D: 2.600. Yêu cầu ghi vào bảng thống kê số liệu trên.

Ở đây dễ dàng thấy tiêu thức A là Bưu điện tỉnh. Nó có 04 biểu hiện đó là Bưu điện tỉnh A (A_1), Bưu điện tỉnh B (A_2), Bưu điện tỉnh C (A_3) và Bưu điện tỉnh D (A_4). Tần số tương ứng ở đây chính là doanh thu của mỗi Bưu điện tỉnh.

Bảng 2.2

Bưu điện	Doanh thu f_i (Triệu đồng)	Tần suất d_i
Tỉnh A	4.000	0,20
Tỉnh B	2.400	0,12
Tỉnh C	11.000	0,55
Tỉnh D	2.600	0,13
Cộng	20.000	1

Trường hợp thứ hai: Xét đồng thời biểu hiện của 02 tiêu thức liên quan.

Giả sử ngoài tiêu thức A, còn xét đồng thời tiêu thức B với những biểu hiện B_j ($j = 1, 2, \dots, l$). Khi đó bảng thống kê có dạng sau:

Bảng 2.3

Tiêu thức B Tiêu thức A	B ₁	B ₂	...	B _j	...	B _i	f _i ^A
A ₁	f ₁₁	f ₁₂	...	f _{1j}	...	f _{1i}	f ₁ ^A
A ₂	f ₂₁	f ₂₂	...	f _{2j}	...	f _{2i}	f ₂ ^A
...	...	...	...	...	...	...	...
A _i	f _{i1}	f _{i2}	...	f _{ij}	...	f _{ii}	f _i ^A
...	...	...	...	...	...	...	...
A _k	f _{k1}	f _{k2}	...	f _{kj}	...	f _{ki}	f _k ^A
Cộng	f ₁ ^B	f ₂ ^B	...	f _j ^B	...	f _i ^B	n = $\sum f_i^A = \sum f_j^B$

Trong bảng thống kê này mỗi f_{ij} là số lần quan sát (tần số), kết quả của sự tồn tại đồng thời 02 biểu hiện A_i và B_j. Cột cuối cùng (f_i^A) và hàng cuối cùng (f_j^B) là tổng riêng các tần số ứng với từng biểu hiện của mỗi tiêu thức A và B.

Như vậy, bảng thống kê số liệu của 02 tiêu thức chính là biểu hiện sự kết hợp 02 bảng thống kê (loại một tiêu thức) trong đó có sự bổ sung các tần số kết hợp (f_{ij}).

Ví dụ: Thêm số liệu ở trường hợp 1 với các điều kiện doanh thu theo loại dịch vụ:

- Doanh thu dịch vụ bưu phẩm thường là: 3.500 triệu đồng của Bưu điện tỉnh A

- Doanh thu dịch vụ bưu phẩm ghi số là:

500 triệu đồng của Bưu điện tỉnh A

1.500 triệu đồng của Bưu điện tỉnh B

500 triệu đồng của Bưu điện tỉnh C

- Doanh thu dịch vụ EMS là:

500 triệu đồng của Bưu điện tỉnh B

1.000 triệu đồng của Bưu điện tỉnh C

- Doanh thu điện thoại đường dài là:

400 triệu đồng của Bưu điện tỉnh B

1.500 triệu đồng của Bưu điện tỉnh C

- Doanh thu điện thoại nội hạt là:

8.000 triệu đồng của Bưu điện tỉnh C

- Doanh thu điện thoại quốc tế là:

2.600 triệu đồng của Bưu điện tỉnh D

Nếu chỉ quan tâm đến tổng số doanh thu từng dịch vụ thì tiêu thức là dịch vụ với 06 biểu hiện.

Bảng 2.4

Loại dịch vụ	Doanh thu f_i (Triệu đồng)	Tần suất d_i
Bưu phẩm thường	3.500	0,175
Bưu phẩm ghi số	2.500	0,125
Chuyển phát nhanh	1.500	0,075
Điện thoại đường dài	1.900	0,095
Điện thoại nội hạt	8.000	0,400
Điện thoại quốc tế	2.600	0,130
Cộng	20.000	1

Trường hợp xét biểu hiện cả 02 tiêu thức Bưu điện tỉnh và loại dịch vụ, ta có bảng thống kê như sau:

Bảng 2.5

Đơn vị: Triệu đồng

Loại dịch vụ Bưu điện tỉnh	Bưu phẩm			Điện thoại			Cộng theo bưu điện tỉnh
	Thường	Ghi số	EMS	Đường dài	Nội hạt	Quốc tế	
A	3.500	500					4.000
B		1.500	500	400			2.400
C		500	1.000	1.500	8.000		11.000
D						2.600	2.600
Cộng theo loại dịch vụ	3.500	2.500	1.500	1.900	8.000	2.600	20.000

2.1.2. Đối với tiêu thức số lượng

Trường hợp 1: Tiêu thức số lượng với các lượng biến rời rạc

Giá trị của các lượng biến này thường là những số nguyên như số máy móc, số công nhân. Trình bày số liệu bằng bảng thống kê đối với tiêu thức số lượng có các lượng biến rời rạc tương đối đơn giản. Với tập hợp số liệu đã có, chỉ việc đếm xem mỗi giá trị lượng biến xuất hiện bao nhiêu lần, rồi ghi kết quả đó vào bảng thống kê.

Bảng 2.6

Các lượng biến x_i	Tần số f_i	Tần suất d_i
x_1	f_1	d_1
x_2	f_2	d_2
...	...	...
x_i	f_i	d_i
...	...	...
x_k	f_k	d_k
Cộng	$n = \sum f_i$	$\sum d_i = 1$

Trong bảng x_i ($i = 1, 2 \dots k$) - Giá trị các lượng biến thuộc tiêu thức x

f_i ($i = 1, 2 \dots k$) - Tần số tương ứng với mỗi x_i

Cần phân biệt lượng số liệu tức là tổng số lần quan sát (n) với lượng các giá trị (k).

Ví dụ: Khảo sát 50 bưu cục cho thấy doanh thu hàng tháng tính bằng triệu đồng như sau:

30	38	34	25	35	33	27	25	33	30
33	30	35	34	27	25	30	37	28	30
35	28	33	28	30	33	28	27	30	28
33	27	30	35	28	34	30	35	38	34
34	33	34	30	38	33	33	33	34	30

Với mẫu biểu 2.6 ta có thể trình bày số liệu doanh thu như sau:

Bảng 2.7

Doanh thu (triệu đồng)	Số bưu cục	Tần suất
x_i	f_i	d_i
25	3	0,06
27	4	0,08
28	6	0,12
30	11	0,22
33	10	0,20
34	8	0,16
35	5	0,10
38	3	0,06
Cộng	50	1

Trường hợp 2: Tiêu thức số lượng với các lượng biến liên tục

Với loại tiêu thức này, trong một khoảng nào đó các lượng biến có thể lấy những giá trị bất kỳ. Trong thực tế quan sát chỉ thu nhận được các giá trị tách biệt tức là các lượng biến liên tục đã bị rời rạc hóa. Điều này không có nghĩa là việc trình bày số liệu với các lượng biến liên tục hoàn toàn giống như các lượng biến rời rạc (trừ trường hợp số lượng các giá trị tương đối ít). Thông thường hay gặp những trường hợp số liệu quan sát cũng như lượng các giá trị rất lớn. Để cho gọn, phải hệ thống hóa lại thành bảng phân phối ghép tổ, tức là các giá trị sẽ được ghép thành từng tổ. Căn cứ vào giới hạn của các tổ, xác định tần số (tần suất) tương ứng và ghi vào bảng phân phối ghép tổ.

Khi tiến hành phân tổ, nếu số tổ quá nhiều sẽ bị xé lẻ và số đơn vị của tổng thể bị phân tán vào nhiều tổ có tính chất giống nhau. Ngược lại, nếu số tổ quá ít thì nhiều đơn vị có tính chất khác nhau sẽ bị phân vào cùng một tổ làm cho mọi kết luận rút ra sau này kém chính xác. Nếu gọi K là số tổ cần thiết thì có thể xác định bằng công thức xấp xỉ sau:

$$K = 3,322 \lg n + 1$$

Trong đó: $\lg n$ - Lô-ga-rít cơ số 10 của n

n - Tổng số lần quan sát

Sau khi xác định số tổ cần thiết, tính độ lớn của mỗi tổ:

$$h = \frac{X_{\max} - X_{\min}}{K}$$

Trong đó: h - Độ lớn mỗi tổ

$x_{\max}$, $x_{\min}$ - Giá trị lớn nhất và nhỏ nhất trong số tất cả các giá trị các lượng biến x_i mà ta quan sát được.

Việc xác định tần số của mỗi tổ được tiến hành tương tự như trong trường hợp đối với lượng biến rời rạc tức là lần lượt đếm các giá trị lượng biến thuộc từng tổ. Những giá trị nằm vào ranh giới của hai tổ liền nhau tùy theo quy ước mà giá trị đó xếp vào tổ trước hay tổ sau.

2.1.3. Những yêu cầu đối với việc xây dựng bảng thống kê

- Quy mô bảng thống kê không nên quá lớn (quá nhiều tổ hoặc quá nhiều chỉ tiêu giải thích). Khi có nhiều tiêu thức cần phân tổ có nhiều chỉ tiêu giải thích thì nên tách ra xây dựng một số bảng thống kê.

- Các tiêu đề, tiêu mục trong bảng thống kê phải được ghi chính xác, đầy đủ, ngắn gọn và dễ hiểu.

- Các chỉ tiêu giải thích trong bảng thống kê cần được sắp xếp theo thứ tự hợp lý, phù hợp với mục đích nghiên cứu. Các chỉ tiêu có liên hệ với nhau nên sắp xếp gần nhau.

- Các ô trong bảng thống kê dùng để ghi các con số thống kê. Nếu không có số liệu để ghi vào một hoặc một số ô nào đó thì dùng các ký hiệu theo quy ước.

Trong bảng thống kê phải dùng đơn vị tính cụ thể cho từng chỉ tiêu. Nếu tất cả các số trong bảng có cùng đơn vị thì đơn vị tính ghi ở đầu bảng. Nếu các chỉ tiêu có đơn vị tính khác nhau thì đơn vị tính ghi ngay dưới tiêu mục.

Dưới bảng thống kê cần ghi rõ nguồn tài liệu sử dụng và các chi tiết cần thiết mà không thể hiện được trong bảng thống kê.

2.2. TRÌNH BÀY SỐ LIỆU BẰNG ĐỒ THỊ

2.2.1. Ý nghĩa và tác dụng của đồ thị

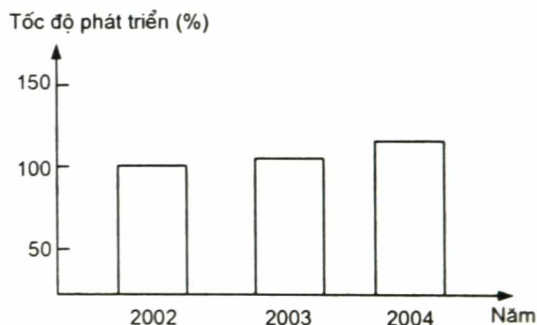
Đồ thị thống kê là các hình vẽ hoặc đường nét hình học dùng để miêu tả có tính chất quy ước các tài liệu thống kê. Ví dụ có tài liệu về tốc độ phát triển doanh thu kinh doanh BCVT của một doanh nghiệp như sau:

Năm 2002 đạt 13.177,80 tỷ đồng = 100%

Năm 2003 đạt 15.294,04 tỷ đồng = 116,59%

Năm 2004 đạt 18.617,15 tỷ đồng = 121,73%

Với tài liệu trên, để dễ thấy, dễ hiểu sự phát triển doanh thu phát sinh, người ta có thể dùng hình vẽ sau:



Trong hình vẽ trên, các cột đứng nói lên sự phát triển của doanh thu phát sinh của doanh nghiệp qua các năm 2002, 2003, 2004. Với cùng một bề rộng, các chiều cao khác nhau của mỗi hình cột cho ta một nhận thức có so sánh cụ thể về sự phát triển của hiện tượng nghiên cứu. Các hình vẽ gọi là biểu đồ thống kê, còn phương pháp dùng các hình vẽ để miêu tả hiện tượng qua các con số thống kê gọi là phương pháp đồ thị thống kê.

Đồ thị thống kê có mấy đặc điểm sau:

- Đồ thị thống kê sử dụng con số kết hợp với hình vẽ, đường nét và màu sắc để trình bày và phân tích các đặc trưng số lượng của hiện tượng. Vì vậy người xem không mất nhiều thời gian đọc con số mà vẫn nhận thức được vấn đề chủ yếu một cách dễ dàng, nhanh chóng.

- Đồ thị thống kê chỉ trình bày một cách khái quát các đặc điểm chủ yếu về bản chất và xu hướng phát triển của các hiện tượng.

Do các đặc điểm nêu trên, đồ thị thống kê có tính phổ cập, có sức hấp dẫn và sinh động, làm cho người hiểu biết ít về thống kê vẫn lĩnh hội được vấn đề chủ yếu một cách dễ dàng.

Phương pháp đồ thị thống kê được ứng dụng rộng rãi trong nghiên cứu, nhằm mục đích hình tượng hóa:

- Sự phát triển của hiện tượng qua thời gian;
- Kết cấu và biến động của kết cấu qua hiện tượng;
- Trình độ phổ biến của hiện tượng;
- Sự so sánh giữa các mức độ của hiện tượng;
- Mối liên hệ giữa các hiện tượng ;
- Tình hình thực hiện kế hoạch.

Ngoài ra, đồ thị thống kê còn được coi là một phương tiện tuyên truyền, một công cụ dùng để biểu thị các kết quả sản xuất.

2.2.2. Các loại đồ thị thống kê

Căn cứ theo nội dung phản ánh, người ta có thể phân chia đồ thị thống kê thành các loại sau:

- Đồ thị so sánh;

- Đồ thị phát triển;
- Đồ thị kết cấu;
- Đồ thị hoàn thành kế hoạch;
- Đồ thị liên hệ;
- Đồ thị phân phối.

Căn cứ vào hình thức biểu hiện, người ta có thể phân chia thành:

- Biểu đồ hình cột;
- Biểu đồ tượng hình;
- Biểu đồ diện tích;
- Đồ thị đường gấp khúc.

Tùy theo nội dung cần phản ánh và trường hợp cụ thể để lựa chọn các hình thức biểu hiện phù hợp.

2.2.3. Yêu cầu đối với việc xây dựng đồ thị thống kê

Một đồ thị thống kê phải đảm bảo các yêu cầu là chính xác, dễ xem, dễ hiểu; ngoài ra còn phải thể hiện tính thẩm mỹ của đồ thị, cho nên khi xây dựng đồ thị thống kê cần thực hiện các yêu cầu sau:

- Lựa chọn loại đồ thị cho phù hợp với nội dung, tính chất các số liệu cần diễn đạt. Để đảm bảo những yêu cầu trên, cần chú ý đến các yếu tố của đồ thị, quy mô, các ký hiệu hình học hoặc các hình vẽ, hệ tọa độ, thang và tỷ lệ xích, phần giải thích.

- Xác định quy mô đồ thị cho thích hợp. Quy mô của đồ thị được quyết định bởi chiều dài, chiều cao và quan hệ tỷ lệ giữa 02 chiều đó. Quy mô của đồ thị lớn hay nhỏ còn phải căn cứ vào mục đích sử dụng. Quan hệ tỷ lệ giữa chiều cao và chiều dài của đồ thị, thông thường được dùng từ 1 : 1,33 đến 1 : 1,50.

- Các ký hiệu hình học hoặc hình vẽ quyết định dáng của đồ thị. Các ký hiệu hình học có nhiều loại như: các chấm, các đường thẳng hoặc cong, các hình cột, hình vuông, hình chữ nhật, hình tròn v.v... Các hình vẽ khác trên đồ thị cũng có thể thay đổi nhiều loại tùy theo tính chất của hiện tượng nghiên cứu. Việc lựa chọn các ký hiệu hình học hoặc hình vẽ của đồ thị là vấn đề quan trọng, vì mỗi hình có khả năng diễn tả riêng, như khi cần biểu hiện kết cấu thành phần thời gian quay vòng toa xe có thể vẽ các hình cột hoặc các hình tròn (có chia thành các hình quạt).

- Hệ tọa độ giúp cho việc xác định chính xác vị trí các ký hiệu hình học trên đồ thị. Các đồ thị thống kê thường dùng hệ tọa độ vuông góc. Trên hệ tọa độ vuông góc, trục hoành thường được dùng để biểu thị thời gian, trục tung biểu thị trị số của chỉ tiêu. Trong trường hợp phân tích mối liên hệ giữa hai tiêu thức, thì tiêu thức nguyên nhân được ghi trên trục hoành, tiêu thức kết quả được ghi trên trục tung.

- Thang và tỷ lệ xích giúp cho việc tính chuyển các đại lượng lên đồ thị theo các khoảng cách thích hợp. Người ta thường dùng các thang đường thẳng, được phân theo các trục tọa độ, cũng có khi dùng thang đường cong như thang tròn (ở đồ thị hình tròn) được chia thành 360°. Các thang tỷ lệ có thể có khoảng cách bằng nhau hoặc không bằng nhau. Các thang tỷ lệ có các khoảng cách không bằng nhau chỉ dùng để biểu hiện các tốc độ khi khoảng biến thiên của các mức độ quá lớn mà người ta chỉ chú ý đến biến động tương đối của chúng.

- Phải giải thích tên đồ thị, các con số và ghi chú dọc theo thang tỷ lệ, các con số bên cạnh từng bộ phận của đồ thị, giải thích các ký hiệu quy ước... cần được ghi rõ, ngắn gọn, dễ hiểu.

Chương 3

CÁC MỨC ĐỘ CỦA HIỆN TƯỢNG THỐNG KÊ

3.1. SỐ TUYỆT ĐỐI

3.1.1. Khái niệm và ý nghĩa của số tuyệt đối

Trong thống kê, số tuyệt đối biểu hiện quy mô, khối lượng hiện tượng trong điều kiện thời gian và không gian cụ thể.

Số tuyệt đối trong thống kê có thể biểu hiện số đơn vị của tổng thể hay bộ phận (Số doanh nghiệp, số công nhân) hoặc các trị số của một tiêu thức hay chỉ tiêu thống kê nào đó (khối lượng sản phẩm dịch vụ, tổng chi phí sản xuất BCVT)...

Số tuyệt đối có ý nghĩa quan trọng đối với công tác quản lý kinh tế như sau:

- Thông qua số tuyệt đối có thể nhận thức được cụ thể về quy mô, khối lượng của hiện tượng nghiên cứu, khả năng tiềm tàng và kết quả đạt được về sự phát triển kinh tế xã hội.
- Số tuyệt đối là cơ sở đầu tiên để phân tích thống kê và là cơ sở để tính các loại chỉ tiêu thống kê khác như số tương đối, số bình quân.
- Là căn cứ để xây dựng kế hoạch và chỉ đạo thực hiện kế hoạch.

3.1.2. Đặc điểm của số tuyệt đối

Số tuyệt đối trong thống kê bao giờ cũng phải mang một nội dung kinh tế nào đó trong thời gian và không gian cụ thể, có

nghĩa là số tuyệt đối trong thống kê phải phản ánh được nội dung gì? Của đơn vị nào? Khi nào?

Số tuyệt đối trong thống kê phải được xác định qua điều tra và tổng hợp thống kê, nó không phải là con số lựa chọn tùy ý, nó được xác định thông qua một hoặc một số phép tính toán.

3.1.3. Phân loại số tuyệt đối

Tùy theo tính chất của hiện tượng nghiên cứu, người ta có thể phân biệt hai loại số tuyệt đối sau:

3.1.3.1. Số tuyệt đối thời kỳ

Số tuyệt đối thời kỳ là số tuyệt đối phản ánh quy mô, khối lượng của hiện tượng trong một độ dài thời gian nhất định. Nó được hình thành thông qua sự tích lũy về lượng của hiện tượng trong suốt thời gian nghiên cứu.

Ví dụ: Doanh thu kinh doanh các dịch vụ BCVT năm 2004 là 30.662,77 tỷ đồng; tổng số máy điện thoại trên mạng hiện có đến 31 tháng 12 năm 2004 là 9.906.091 máy.

Các số tuyệt đối thời kỳ của cùng một chỉ tiêu có thể cộng lại với nhau để có trị số của thời kỳ dài hơn. Thời kỳ càng dài, trị số của chỉ tiêu càng lớn.

3.1.3.2. Số tuyệt đối thời điểm

Số tuyệt đối thời điểm là số tuyệt đối phản ánh mặt lượng của hiện tượng nghiên cứu tại một thời điểm nhất định. Với số tuyệt đối thời điểm thì không thể cộng lại với nhau vì khi cộng lại chúng không phản ánh được nội dung kinh tế nào.

Ví dụ: Không thể cộng số lao động hiện có tại hai thời điểm là 01/4/2003 và 01/4/2004.

Số tuyệt đối thời điểm chỉ phản ánh tình hình của hiện tượng tại một thời điểm nào đó, trước và sau thời điểm trạng thái của hiện tượng này có thể khác nhau. Muốn có số tuyệt đối thời điểm chính xác phải quy định thời điểm hợp lý và phải tổ chức điều tra kịp thời.

3.1.4. Đơn vị tính số tuyệt đối

Trong thống kê, các số tuyệt đối đều có đơn vị tính cụ thể để biểu thị nội dung của hiện tượng nghiên cứu. Tùy theo tính chất của hiện tượng và mục đích nghiên cứu, người ta có thể sử dụng các đơn vị tính khác nhau như sau:

Đơn vị tự nhiên: Là đơn vị tính phù hợp với đặc điểm vật lý của hiện tượng như cái, con, chiếc... hoặc thời gian (phút, giờ, ngày...). Trong nhiều trường hợp phải dùng đơn vị kép như mật độ điện thoại (máy/100 dân).

Đơn vị thời gian: Thường dùng để tính lượng lao động hao phí sản xuất ra những sản phẩm dịch vụ không thể tổng hợp hoặc so sánh với nhau bằng đơn vị tính toán khác, hoặc những sản phẩm dịch vụ phức tạp do nhiều người thực hiện qua nhiều giai đoạn khác nhau.

Đơn vị tiền tệ (VND, USD, EURO,...): được sử dụng rộng rãi nhất để biểu thị giá trị của sản phẩm dịch vụ. Nó giúp cho việc tổng hợp và so sánh nhiều loại sản phẩm có giá trị sử dụng và đo lường khác nhau. Tuy nhiên, do giá cả luôn thay đổi nên đơn vị tiền tệ không có tính chất so sánh được qua thời gian. Để khắc

phục nhược điểm này, người ta phải dùng “giá cố định” ở thời gian nào đó.

3.2. SỐ TƯƠNG ĐỐI

3.2.1. Khái niệm và ý nghĩa số tương đối

3.2.1.1. Khái niệm

Số tương đối trong thống kê là một loại chỉ tiêu biểu hiện quan hệ so sánh giữa hai mức độ của hiện tượng nghiên cứu cùng loại nhưng khác nhau về thời gian hoặc không gian, hoặc biểu diễn quan hệ tỷ lệ giữa hai chỉ tiêu thống kê khác loại nhưng có liên quan đến nhau.

Ví dụ: Doanh thu khối hạch toán phụ thuộc Tổng Công ty BCVT Việt Nam năm 2004 so với năm 2003 bằng 122,66%. Cũng có thể so sánh các mức độ khác loại nhưng có liên quan với nhau như mật độ điện thoại, số dân phục vụ của một bưu cục...

3.2.1.2. Ý nghĩa

Trong thống kê, số tương đối có ý nghĩa quan trọng. Nó là một trong những chỉ tiêu phân tích thống kê. Nếu như số tuyệt đối chỉ khái quát được về quy mô khối lượng của hiện tượng nghiên cứu, thì số tương đối cho phép phân tích đặc điểm của hiện tượng, nghiên cứu các hiện tượng trong mối quan hệ so sánh với nhau.

Trong thống kê, số tương đối phản ánh các mặt sau:

- Số tương đối cho biết kết cấu của hiện tượng;
- Biểu hiện mối quan hệ giữa hai chỉ tiêu;

- Biểu hiện trình độ phát triển của hiện tượng (so sánh thực hiện giữa các năm);
- Trình độ phổ biến của hiện tượng (bán kính phục vụ, mật độ điện thoại...);
- Giúp ta kiểm tra đánh giá tình hình thực hiện kế hoạch;
- Đi sâu phân tích so sánh đặc điểm của hiện tượng và giữ bí mật số tuyệt đối.

3.2.2. Đặc điểm của số tương đối

Trong thống kê số tương đối không phải là con số trực tiếp thu được qua điều tra mà là kết quả so sánh giữa hai số tuyệt đối đã có. Vì vậy mỗi số tương đối phải có gốc để so sánh (gốc so sánh gọi là chỉ tiêu gốc). Tùy theo mục đích nghiên cứu mà người ta chọn gốc so sánh cho phù hợp.

Ví dụ, để biểu hiện sự phát triển của hiện tượng theo thời gian thì gốc là mức độ kỳ trước; để biểu hiện kết cấu thì gốc là mức độ tổng thể, để biểu hiện mức độ hoàn thành kế hoạch thì gốc là mức độ kế hoạch...

Trong thống kê số tương đối có thể biểu thị bằng lần hoặc phần trăm (%), phần nghìn (‰).

3.2.3. Các loại số tương đối

Căn cứ vào nội dung kinh tế mà số tương đối phản ánh thì số tương đối được chia làm 06 loại như sau: Số tương đối động thái, số tương đối hoàn thành kế hoạch, số tương đối nhiệm vụ kế hoạch, số tương đối kết cấu, số tương đối so sánh, số tương đối cường độ.

3.2.3.1. Số tương đối động thái (tốc độ phát triển, chỉ số phát triển)

Số tương đối động thái biểu hiện sự biến động về mức độ của hiện tượng nghiên cứu qua thời gian. Nó được xác định bằng cách so sánh hai mức độ của chỉ tiêu cùng loại nhưng khác nhau về thời gian, được biểu hiện bằng số lần hoặc số phần trăm.

$$t_{dt} = \frac{y_1}{y_0} \text{ (lần)}, \text{ hoặc } t_{dt} = \frac{y_1}{y_0} 100 \text{ (\%)}$$

Trong đó: y_1 - Mức độ của hiện tượng kỳ nghiên cứu (kỳ báo cáo);

y_0 - Mức độ của hiện tượng kỳ gốc;

t_{dt} - Số tương đối động thái.

Muốn có số tương đối động thái chính xác cần đảm bảo tính chất so sánh được giữa các mức độ của kỳ báo cáo và kỳ gốc, tức là phải đảm bảo giống nhau về nội dung kinh tế, về phương pháp tính và đơn vị tính, về phạm vi và độ dài thời gian mà mức độ phản ánh.

3.2.3.2. Số tương đối kế hoạch

Số tương đối kế hoạch được dùng để lập và kiểm tra tình hình thực hiện kế hoạch về một chỉ tiêu nào đó, có hai loại số tương đối kế hoạch:

- Số tương đối hoàn thành kế hoạch biểu hiện quan hệ tỷ lệ giữa mức độ thực tế đạt được trong kỳ so với mức độ kế hoạch đã đề ra của chỉ tiêu kinh tế nào đó.

$$t_{htkh} = \frac{y_1}{y_{kh}} \text{ hoặc } t_{htkh} = \frac{y_1}{y_{kh}} \times 100$$

Trong đó: y_{kh} - Mức độ kế hoạch;

y_l - Mức độ thực tế của kỳ báo cáo;

t_{hkh} - Số tương đối kế hoạch.

- Số tương đối nhiệm vụ kế hoạch biểu hiện mối quan hệ tỷ lệ giữa mức độ cần đạt tới của chỉ tiêu kinh tế nào đó trong kỳ kế hoạch so với mức độ thực tế của chỉ tiêu đó ở kỳ gốc. Như vậy số tương đối nhiệm vụ kế hoạch phản ánh mục tiêu cần đạt tới của đơn vị.

$$t_{nvkh} = \frac{y_{kh}}{y_o} \text{ hoặc } t_{nvkh} = \frac{y_{kh}}{y_o} \times 100$$

Trong đó: t_{nvkh} - Số tương đối nhiệm vụ kế hoạch;

Giữa số tương đối động thái, số tương đối hoàn thành kế hoạch, số tương đối nhiệm vụ kế hoạch của cùng một chỉ tiêu, cùng một thời gian có quan hệ như sau:

$$t_{dt} = t_{nvkh} \times t_{hkh}$$

Quan hệ này được vận dụng để tính mức độ chưa biết khi đã biết các mức độ kia.

3.2.3.3. Số tương đối kết cấu

Số tương đối kết cấu phản ánh tỷ trọng của mỗi bộ phận cấu thành trong một tổng thể. Qua chỉ tiêu này có thể phân tích được đặc điểm cấu thành của hiện tượng.

Số tương đối kết cấu được tính bằng cách so sánh trị số tuyệt đối của chỉ tiêu của từng bộ phận so với trị số tuyệt đối của chỉ tiêu của cả tổng thể. Nó thường được biểu hiện bằng lần hoặc số phần trăm.

$$t_{kc} = \frac{y_{bp}}{y_{tt}} \text{ hoặc } t_{kc} = \frac{y_{bp}}{y_{tt}} \times 100$$

Trong đó: y_{bp} - Trị số tuyệt đối của chỉ tiêu của bộ phận;

y_{tt} - Trị số tuyệt đối của chỉ tiêu của tổng thể;

t_{kc} - Số tương đối kết cấu.

Muốn tính số tương đối kết cấu chính xác phải phân biệt được các bộ phận có tính chất khác nhau trong tổng thể nghiên cứu, có nghĩa là phải phân tổ chính xác, tức là phải dựa vào phương pháp phân tổ thống kê.

3.2.3.4. Số tương đối so sánh

Số tương đối so sánh biểu hiện quan hệ so sánh giữa các hiện tượng cùng loại khác nhau về không gian. Số tương đối so sánh còn biểu hiện sự so sánh giữa các bộ phận trong cùng một tổng thể, khi so sánh người ta lấy một bộ phận nào đó làm gốc rồi đem các bộ phận khác so sánh với nó.

3.2.3.5. Số tương đối cường độ

Biểu hiện trình độ phổ biến của hiện tượng nghiên cứu trong điều kiện lịch sử nhất định, nó được xác định bằng cách so sánh chỉ tiêu của hai hiện tượng khác nhau nhưng có liên quan với nhau. Mức độ của hiện tượng mà ta cần nghiên cứu trình độ phổ biến được đặt ở tử số, còn mức độ của hiện tượng có liên quan được đặt ở mẫu số. Hình thức biểu hiện của số tương đối cường độ là đơn vị kép do đơn vị tính của tử số và mẫu số hợp thành.

Ví dụ: Mật độ điện thoại trên 100 dân

= Số máy điện thoại $\times$ 100/Tổng dân số (máy/100 dân)

3.3. SỐ BÌNH QUÂN

3.3.1. Khái niệm số bình quân

Số bình quân trong thống kê là loại chỉ tiêu biểu hiện mức độ điển hình theo một tiêu thức hoặc chỉ tiêu thống kê nào đó của một tổng thể bao gồm nhiều đơn vị cùng loại.

Một tổng thể thống kê thường bao gồm nhiều đơn vị, các đơn vị này cơ bản là giống nhau, nhưng biểu hiện cụ thể về mặt lượng theo các tiêu thức nào đó thường chênh lệch nhau. Những chênh lệch này do nhiều yếu tố, ngoài các nguyên nhân chung quyết định đặc điểm cơ bản của hiện tượng, còn có những nguyên nhân riêng ảnh hưởng tới từng đơn vị. Khi nghiên cứu tổng thể thống kê không chỉ nêu các đặc điểm riêng biệt của từng đơn vị, mà cần tìm một chỉ tiêu có khả năng mô tả một cách khái quát đặc điểm chung, điển hình nhất của hiện tượng. Mức độ đó chính là số bình quân.

Ví dụ khi so sánh thu nhập của lao động trong các đơn vị Bưu điện, không thể lấy tiền lương của bất kỳ một nhóm công nhân nào đó làm đại diện. Bởi vì tiền lương của các nhóm công nhân là khác nhau do trình độ, do điều kiện làm việc, sức khỏe... Do đó có thể lấy tiền lương bình quân của tất cả các công nhân trong một đơn vị để so sánh về mức độ thu nhập của các đơn vị với nhau.

3.3.2. Ý nghĩa của số bình quân

Số bình quân có tính tổng hợp và khái quát cao, chỉ cần một trị số là số bình quân có thể nêu lên mức độ chung nhất, phổ biến nhất của tiêu thức nghiên cứu. Nó không kể đến chênh lệch thực tế giữa các đơn vị tổng thể. Số bình quân không biểu hiện mức độ

cá biệt mà biểu hiện mức độ tính chung nhất của các đơn vị tổng thể.

Do số bình quân biểu hiện đặc điểm chung của cả tổng thể nghiên cứu cho nên những nét riêng biệt có tính chất ngẫu nhiên của từng đơn vị cá biệt bị loại trừ, tức là số bình quân đã san bằng mọi chênh lệch giữa các đơn vị tổng thể về lượng biến của tiêu thức nghiên cứu. Sự san bằng này chỉ có ý nghĩa lớn khi tính cho một số khá nhiều đơn vị.

Số bình quân có vị trí quan trọng trong lý luận và trong công tác thực tế. Nó được sử dụng để:

- Nêu lên đặc điểm điển hình của hiện tượng kinh tế xã hội số lớn trong điều kiện thời gian và địa điểm cụ thể;

- Dùng để so sánh các hiện tượng không cùng quy mô (có thể so sánh kết quả của hoạt động sản xuất kinh doanh của 02 đơn vị A, B, có thể 02 đơn vị này có số vốn khác nhau, có số lao động khác nhau nhưng có thể so sánh năng suất lao động bình quân của 02 đơn vị);

- Có thể dùng để nghiên cứu sự biến động của hiện tượng theo thời gian, nhằm thấy được xu hướng phát triển cơ bản của hiện tượng số lớn, nghĩa là của đại bộ phận các đơn vị.

- Số bình quân được dùng trong công tác lập kế hoạch, phân tích thống kê.

3.3.3. Các loại số bình quân trong thống kê

Trong thống kê có rất nhiều loại số bình quân và khi sử dụng loại nào cần căn cứ vào đặc điểm nguồn tài liệu sẵn có, mục đích

nghiên cứu và ý nghĩa kinh tế xã hội của số bình quân. Số bình quân được phân thành hai nhóm:

- Nhóm 1: Các số bình quân được tính từ tất cả các lượng biến bằng cách chia đều tổng các lượng biến cho số đơn vị có lượng biến đó. Nhóm này bao gồm số trung bình cộng, trung bình nhân, trung bình điều hòa.

- Nhóm 2: Số bình quân chuyên dùng để mô tả đặc trưng phân phối các đơn vị của tổng thể theo trị số tiêu thức nghiên cứu. Nhóm này gồm có mốt và trung vị.

3.3.3.1. Số bình quân số học (Trung bình cộng)

Số bình quân số học còn gọi là số trung bình cộng được xác định bằng cách đem chia tổng lượng biến của tiêu thức cho tổng số đơn vị tổng thể.

* Số trung bình cộng giản đơn

Nó được sử dụng trong trường hợp mỗi lượng biến chỉ xuất hiện một lần, công thức tính như sau:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

Trong đó: x_i - Lượng biến của tiêu thức của đơn vị thứ i ;

$\bar{x}$ - Số trung bình của tổng thể;

n - Số đơn vị tổng thể.

* Số trung bình cộng gia quyền (số trung bình số học có trọng số, số bình quân gia quyền)

Khi mỗi lượng biến x_i xuất hiện nhiều lần, tức là có tần số (f_i) khác nhau, muốn xác định số trung bình cộng cần nhân mỗi

lượng biến x_i với tần số f_i tương ứng, sau đó cộng lại và đem chia cho số đơn vị tổng thể:

$$\bar{X} = \frac{x_1 f_1 + x_2 f_2 + x_3 f_3 + \dots + x_m f_m}{f_1 + f_2 + \dots + f_m} = \frac{\sum_{i=1}^m x_i f_i}{\sum_{i=1}^m f_i}$$

Trong đó: $\sum_{i=1}^m f_i$ - Số đơn vị tổng thể

x_i - Là lượng biến i ;

f_i - Tần số của tổ i ;

m - Là số tổ;

$x_i f_i$ - Gọi là gia quyền.

*** Các dạng đặc biệt của số trung bình cộng gia quyền**

- Tính số trung bình cộng gia quyền từ dãy số lượng biến có khoảng cách tổ.

Trong trường hợp mỗi tổ có một phạm vi lượng biến, cần có lượng biến đại diện để làm căn cứ tính toán và lượng biến đại diện là các trị số giữa:

x'_i là trị số giữa của tổ thứ i

$$x'_i = \frac{x_{i \min} + x_{i \max}}{2}$$

$x_{i \min}$, $x_{i \max}$ là giới hạn dưới, giới hạn trên của tổ i

$$\bar{X} = \frac{\sum_{i=1}^m x'_i f_i}{\sum_{i=1}^m f_i}$$

- Tính số trung bình cộng chung của các số trung bình cộng tổ.

Trong trường hợp này, số trung bình cộng chung sẽ là số trung bình cộng gia quyền của các số trung bình cộng tổ, trong đó quyền số là số đơn vị của mỗi tổ:

$$\bar{x} = \frac{\sum_{i=1}^m \bar{x}_i f_i}{\sum_{i=1}^m f_i}$$

Trong đó: $\bar{x}_i$ - Là số trung bình cộng của tiêu thức của tổ thứ i ;

f_i - Là tần số của tổ thứ i ;

m - Là số tổ.

- Tính số trung bình cộng gia quyền khi biết tỷ trọng và số đơn vị tổng thể của mỗi tổ:

d_i - Tỷ trọng của tổ thứ i trong tổng thể: $d_i = \frac{f_i}{\sum f_i}$

$$\bar{x} = \sum_{i=1}^m x_i d_i$$

Trong đó: x_i - Là lượng biến của tiêu thức của tổ có tỷ trọng d_i

Để tính số bình quân cộng nhanh chóng cần biết các tính chất của chúng.

Tính chất 1: Nếu đem tất cả các lượng biến x_i cộng thêm (hoặc trừ đi) một số a bất kỳ thì số bình quân tính được cũng sẽ tăng thêm (hoặc giảm đi) một trị số bằng a .

$$\frac{\sum (x_i \pm a) \times f_i}{\sum f_i} = \bar{x} \pm a$$

Tính chất 2: Nếu đem nhân (hoặc chia) tất cả các lượng biến x_i với cùng một số d bất kỳ thì số bình quân tính được cũng sẽ lớn hơn (hoặc nhỏ hơn) d lần.

$$\frac{\sum (x_i \times d) \times f_i}{\sum f_i} = d \times \bar{x}$$

$$\frac{\sum (x_i : d) \times f_i}{\sum f_i} = \bar{x} : d$$

Tính chất 3: Nếu đem nhân (hoặc chia) tất cả các quyền số f_i với cùng một số m bất kỳ thì số bình quân tính được vẫn không thay đổi.

$$\frac{\sum x_i (m \times f_i)}{\sum m \times f_i} = \bar{x}$$

$$\frac{\sum x_i (f_i : m)}{\sum (f_i : m)} = \bar{x}$$

Từ đây có thể suy ra:

- Nếu quyền số của tất cả các lượng biến bằng nhau thì số bình quân cộng gia quyền sẽ bằng số bình quân cộng giản đơn.

$$f_1 = f_2 = \dots = f_n = a$$

$$\frac{\sum x_i f_i}{\sum f_i} = \frac{\sum x_i}{n} = \bar{x}$$

- Khi tính số bình quân cộng, có thể dùng các quyền số là số tương đối tức là tỷ trọng mỗi tổ chiếm trong tổng thể.

$$\begin{aligned}
 \frac{\sum x_i f_i}{\sum f_i} &= \frac{x_1 f_1 + x_2 f_2 + x_3 f_3 + \dots + x_m f_m}{\sum f_i} \\
 &= x_1 \frac{f_1}{\sum f_i} + x_2 \frac{f_2}{\sum f_i} + \dots + x_m \frac{f_m}{\sum f_i} \\
 &= x_1 d_1 + x_2 d_2 + \dots + x_m d_m
 \end{aligned}$$

Tính chất 4: Tổng các độ lệch giữa các lượng biến x_i với số bình quân cộng của chúng bao giờ cũng bằng 0.

$$\sum (x_i - \bar{x}) \times f_i = 0$$

Tính chất 5: Trong một dãy số lượng biến mà tất cả các lượng biến đều có trị số như nhau thì số bình quân tính được chính là trị số đó.

Nếu $x_i = c$ với mọi i thì $\bar{x} = c$.

3.3.3.2. Số bình quân điều hòa

Được sử dụng trong trường hợp khi không có số liệu thống kê về số đơn vị tổng thể nhưng có số liệu thống kê về tổng các lượng biến của tiêu thức. Công thức tính:

$$\bar{x} = \frac{\sum_{i=1}^m M_i}{\sum_{i=1}^m \frac{M_i}{x_i}}$$

Trong đó: x_i - Là lượng biến thứ i ;

$M_i = x_i f_i$ - Là tổng lượng biến của tiêu thức của tổ thứ i ;

m - Là số tổ.

Số bình quân tính theo công thức trên gọi là số bình quân điều hòa gia quyền.

Các trường hợp đặc biệt:

- Nếu các tổng lượng biến của tiêu thức có giá trị bằng nhau, nghĩa là $M_1 = M_2 = M_3 \dots = M$ thì:

$$\bar{x} = \frac{mM}{M \sum_{i=1}^m \frac{1}{x_i}} = \frac{m}{\sum_{i=1}^m \frac{1}{x_i}}$$

- Nếu các tổng lượng biến tiêu thức được xác định dưới tỷ lệ kết cấu, nghĩa là không có tài liệu về M_i về số tuyệt đối, mà chỉ biết các tỷ lệ $M_i / \sum M_i$ thì số bình quân điều hòa được xác định như sau:

$$\bar{x} = \frac{\sum M_i}{\sum \frac{M_i}{x_i}} = \frac{1}{\frac{M_1}{\sum M_i} * \frac{1}{x_1} + \frac{M_2}{\sum M_i} * \frac{1}{x_2} + \dots + \frac{M_m}{\sum M_i} * \frac{1}{x_m}} = \frac{1}{\sum_{i=1}^m \frac{d_i}{x_i}}$$

Trong đó: $d_i = \frac{M_i}{\sum M_i}$

Nếu d_i tính bằng % với $\sum d_i = 100$ thì:

$$\bar{x} = \frac{100}{\sum \frac{d_i}{x_i}}$$

3.3.3.3. Số bình quân nhân

Số bình quân nhân được sử dụng trong trường hợp khi các lượng biến có quan hệ tích số với nhau, thường được sử dụng để tính tốc độ phát triển trung bình của chỉ tiêu thống kê nào đó. Có hai cách tính:

- Số bình quân nhân giản đơn:

Được sử dụng trong trường hợp mỗi lượng biến x_i chỉ xuất hiện một lần:

$$\bar{x} = \sqrt[n]{x_1 \times x_2 \times x_3 \dots \times x_n} = \sqrt[n]{\prod_{i=1}^n x_i}$$

Ví dụ: Tính tốc độ phát triển bình quân hàng năm về chỉ tiêu doanh thu của một đơn vị Bưu điện trong giai đoạn từ năm 1999 - 2004. Nếu biết:

+ Tốc độ phát triển về doanh thu:

Năm 1999 so với năm 1998 là 108 %

Năm 2000 so với năm 1999 là 109%

Năm 2001 so với năm 2000 là 106%

Năm 2002 so với năm 2001 là 108%

Năm 2003 so với năm 2002 là 110%

Năm 2004 so với năm 2003 là 109%

+ Tốc độ phát triển trung bình hàng năm về doanh thu được xác định theo công thức:

$$\begin{aligned}\bar{x} &= \sqrt[6]{\prod_{i=1}^6 x_i} = \sqrt[6]{1,08 \times 1,09 \times 1,06 \times 1,08 \times 1,10 \times 1,09} = \\ &= 1,08326 \text{ (lần) hay } 108,326 \text{ \%}.\end{aligned}$$

- Số bình quân nhân gia quyền:

Được dùng trong trường hợp mỗi lượng biến x_i xuất hiện nhiều lần, tức là có tần số f_i khác nhau:

$$\bar{x} = \sum f_i \sqrt{\prod_1^m x_i^{f_i}}$$

Trong đó: x_i - Là lượng biến thứ i ;

f_i - Là tần số của tổ thứ i ;

m - Là số tổ.

3.3.3.4. Một

a) Khái niệm

Một là biểu hiện của tiêu thức được gặp nhiều nhất trong tổng thể hay trong một dãy số phân phối. Như vậy, trong dãy số lượng biến *một* là lượng biến có tần số lớn nhất. Trị số của *một* không phụ thuộc vào lượng biến của tiêu thức mà phụ thuộc vào sự phân phối các đơn vị trong tổng thể.

Ví dụ: Có tài liệu về số tiếng của 50 bức điện báo như sau:

Bảng 3.1

Số tiếng (x_i)	Số bức điện báo (f_i)
25	2
27	5
28	6
30	10
33	11
34	9
35	6
38	1
Cộng	50

Theo định nghĩa trên có thể xác định *một* là 33 tiếng vì lượng biến này có tần số lớn nhất ($f_{\max} = 11$)

b) Phương pháp xác định *mốt*

Phân biệt các trường hợp sau:

- Đối với dãy số lượng biến không có khoảng cách tổ:
Trường hợp này *mốt* là lượng biến có tần số lớn nhất.

- Đối với dãy số có khoảng cách tổ đều nhau: Muốn tìm *mốt* trước hết phải tìm tổ chứa *mốt*, tổ chứa *mốt* là tổ có tần số lớn nhất, sau đó tính trị số gần đúng của *mốt* theo công thức sau:

$$M_0 = x_{M_{0min}} + h_{M_0} \times \frac{f_{M_0} - f_{M_{0-1}}}{(f_{M_0} - f_{M_{0-1}}) + (f_{M_0} - f_{M_{0+1}})}$$

Trong đó: $x_{M_{0min}}$ - Là giới hạn dưới của tổ chứa *mốt*;

h_{M_0} - Là khoảng cách của tổ có chứa *mốt*;

$f_{M_0}, f_{M_{0-1}}, f_{M_{0+1}}$ - Là tần số của tổ chứa *mốt*, tổ đứng trước tổ chứa *mốt*, và tổ đứng sau tổ chứa *mốt*.

Ví dụ có tài liệu khảo sát 100 đơn vị Bưu điện về tình hình thực hiện doanh thu như sau:

Bảng 3.2

% thực hiện doanh thu	Số đơn vị
75 - 85	6
85 - 95	8
95 - 105	20
105 - 115	24
115 - 125	18
125 - 135	11
135 - 145	7
145 - 155	6
Cộng	100

Tổ có % thực hiện doanh thu từ 105 - 115% là tổ chứa *mốt* vì tổ này có tần số lớn nhất. Thay các số liệu vào công thức ta tính được *mốt*.

$$M_0 = 105 + 10 \times \frac{24 - 20}{(24 - 20) + (24 - 18)} = 109\%$$

- Trường hợp dãy số lượng biến có khoảng cách tổ không đều nhau, tổ chứa *mốt* là tổ có mật độ phân phối lớn nhất, tức là xung quanh lượng biến ấy chứa nhiều tần số nhất. Mật độ phân phối được xác định bằng cách đem chia tần số cho khoảng cách tổ tương ứng:

$$d_i = f_i/h_i$$

Trong đó: d_i - Mật độ phân phối tổ thứ i ;

f_i - Tần số của tổ thứ i ;

h_i - Khoảng cách tổ i .

Khi đó *mốt* được xác định theo công thức:

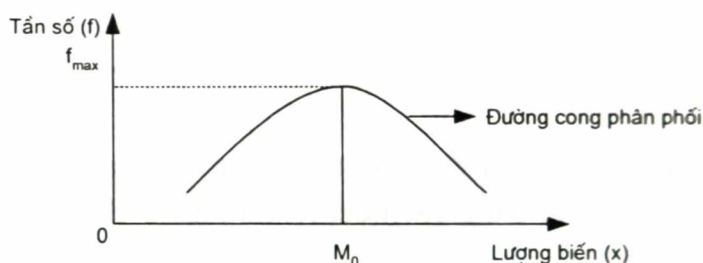
$$M_0 = x_{M_0 \min} + h_{M_0} \times \frac{d_{M_0} - d_{M_0-1}}{(d_{M_0} - d_{M_0-1}) + (d_{M_0} - d_{M_0+1})}$$

c) Ý nghĩa của *mốt*

Mốt biểu hiện mức độ phổ biến nhất của hiện tượng nghiên cứu, đồng thời bản thân nó lại không san bằng, bù trừ chênh lệch giữa các lượng biến, cho nên *mốt* là chỉ tiêu có thể dùng bổ sung hoặc thay thế cho việc tính số trung bình khi cần thiết, trong trường hợp tính số trung bình gặp khó khăn.

Ngoài ra, *mốt* còn là một trong những mức độ để nêu lên đặc trưng phân phối của dãy số như nếu tính được số trung bình, *mốt*, trung vị có giá trị bằng nhau thì dãy số có phân phối chuẩn.

Trên đồ thị *mốt* được xác định như sau:



Trên đồ thị *mốt* được biểu hiện ở điểm trên trục hoành tương ứng với điểm có tung độ lớn nhất thuộc đường cong phân phối.

3.3.3.5. Số trung vị

a) Khái niệm

Số trung vị là lượng biến của tiêu thức của đơn vị đứng ở vị trí giữa trong dãy số lượng biến, chia dãy số lượng biến thành hai phần (phần trên và phần dưới số trung vị), mỗi phần có cùng một số đơn vị tổng thể bằng nhau.

b) Phương pháp tính số trung vị

- Đối với dãy số lượng biến không có khoảng cách tổ

Có hai trường hợp:

+ Nếu số đơn vị của tổng thể lẻ: $n = 2k + 1$, nghĩa là dãy số lượng biến:

$$X_1, X_2, \dots, X_k, X_{k+1}, \dots, X_{2k+1}$$

Thì số trung vị sẽ là lượng biến của đơn vị đứng ở vị trí thứ $k+1$ tức là lượng biến x_{k+1} . Ký hiệu số trung vị là M_c . Ta có:

$$M_c = x_{k+1}$$

+ Nếu số đơn vị tổng thể chẵn ($n = 2k$) nghĩa là dãy số lượng biến:

$$x_1, x_2, \dots, x_k, x_{k+1}, \dots, x_{2k}$$

Thì số trung vị được xác định căn cứ vào lượng biến của hai đơn vị đứng ở vị trí giữa nhất (đơn vị thứ k và $k+1$)

$$M_c = \frac{x_k + x_{k+1}}{2}$$

- Đối với dãy số lượng biến có khoảng cách tổ

Muốn xác định số trung vị đầu tiên cần xác định tổ chứa trung vị, tức là tổ có chứa lượng biến của đơn vị ở vị trí giữa trong tổng số các đơn vị. Tổ có chứa số trung vị - đó là tổ đầu tiên có tần số tích lũy lớn hơn hoặc bằng nửa số đơn vị tổng thể, tức là tổ có tần số tích lũy bằng $(\sum f_i + 1)/2$. Sau đó tính giá trị gần đúng của số trung vị theo công thức sau:

$$M_e = x_{M_{e \min}} + h_{M_e} \times \frac{\sum \frac{f_i}{2} - S_{M_{e-1}}}{f_{M_e}}$$

Trong đó: $x_{M_{e \min}}$ - Là giới hạn dưới của tổ có số trung vị;

h_{M_e} - Khoảng cách tổ có chứa số trung vị;

$S_{M_{e-1}}$ - Tần số tích lũy của tổ đứng trước tổ có số trung vị;

f_{M_e} - Tần số của tổ chứa số trung vị.

Ví dụ: Có tài liệu về phân tổ công nhân theo năng suất lao động của một đơn vị Bưu điện như sau:

Bảng 3.3

Năng suất lao động (Triệu đồng/người)	Số công nhân (người)	Tần số tích lũy (S)
20 - 22	5	5
22 - 24	9	14
24 - 26	20	34
26 - 28	10	44
28 - 30	6	50
Cộng	50	-

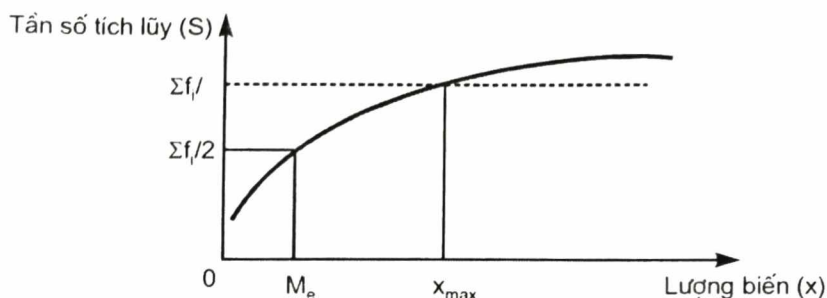
Ta có $(\sum f_i + 1)/2 = (50 + 1)/2 = 25,5$; như vậy số trung vị rơi vào tổ 24 - 26. Do đó số trung vị năng suất lao động là:

$$M_c = 24 + 2 \times \frac{\sum \frac{50}{2} - 14}{20} = 25,1 \text{ (triệu đồng/người)}$$

c) Ý nghĩa

Cũng như một, số trung vị có khả năng nêu lên mức độ điển hình của tổng thể mà không san bằng, bù trừ chênh lệch giữa các lượng biến. Vì vậy có thể bổ sung hoặc thay thế số trung bình cộng khi không có đủ các lượng biến để tính. Việc tính số trung vị cũng thường được dùng để nêu lên đặc trưng của dãy số phân phối.

Trên đồ thị, số trung vị được xác định như sau:



3.4. CÁC CHỈ TIÊU ĐÁNH GIÁ ĐỘ BIẾN THIÊN CỦA TIÊU THỨC

3.4.1. Ý nghĩa nghiên cứu

Khi xét số trung bình, số trung vị, một người ta mới chỉ xét đến tính chất điển hình của toàn bộ tổng thể nghiên cứu mà chưa xét đến từng đơn vị cá biệt và sự chênh lệch thực tế giữa các lượng biến (có khi chênh lệch này là đáng kể hoặc có khi là không đáng kể). Do vậy ngoài việc tính mức độ trung bình, cần phải đánh giá độ biến thiên của tiêu thức nhằm:

- Đánh giá tính chất đại biểu của số trung bình. Nếu độ biến thiên của tiêu thức càng nhiều thì tính chất đại biểu của số trung bình càng ít và ngược lại.

- Trong phân tích các chỉ tiêu hoàn thành kế hoạch và đánh giá độ biến thiên của tiêu thức giúp ta thấy rõ chất lượng công tác và nhịp độ hoàn thành kế hoạch chung và từng bộ phận trong doanh nghiệp.

- Các chỉ tiêu đánh giá độ biến thiên của tiêu thức giúp ta nghiên cứu độ tương quan và tính sai số trong điều tra chọn mẫu (Trình bày ở chương 4, 5).

3.4.2. Các chỉ tiêu đánh giá độ biến thiên của tiêu thức

3.4.1. Khoảng biến thiên (R) hay còn gọi là toàn cự

Khoảng biến thiên là độ chênh lệch giữa lượng biến lớn nhất ($x_{\max}$) và lượng biến nhỏ nhất ($x_{\min}$) của tiêu thức nghiên cứu:

$$R = x_{\max} - x_{\min}$$

Khoảng biến thiên R càng nhỏ thì tính chất đồng đều của tổng thể nghiên cứu càng cao và ngược lại. Chỉ tiêu này giúp xác định nhanh chóng chênh lệch giữa đơn vị tiên tiến và đơn vị lạc hậu. Tuy nhiên nhược điểm của chỉ tiêu này là nó chỉ phụ thuộc vào lượng biến nhỏ nhất và lớn nhất của dãy số, không xét đến các lượng biến khác, cho nên nhiều khi dẫn đến những kết luận không hoàn toàn chính xác.

Ví dụ: Doanh thu tại 02 bưu cục như sau (triệu đồng)

Bưu cục A: 40; 50; 60; 70; 80

Bưu cục B: 58; 59; 60; 61; 62

Doanh thu bình quân của mỗi bưu cục là 60, nhưng thực ra 02 bưu cục này không đồng đều về doanh thu. Doanh thu của bưu cục A chênh lệch nhiều hơn so với bưu cục B. Để đánh giá mức độ biến thiên của doanh thu, qua đó đánh giá tính chất đại biểu của số bình quân, ta tính khoảng biến thiên của 02 bưu cục như sau:

$$R_A = 80 - 40 = 40 \text{ triệu đồng}$$

$$R_B = 62 - 58 = 4 \text{ triệu đồng}$$

Kết quả cho thấy R_A lớn hơn R_B , có nghĩa là độ biến thiên tiêu thức của bưu cục A lớn hơn và vì thế tính chất đại biểu của số bình quân bưu cục A thấp hơn.

3.4.2.2. Độ lệch tuyệt đối trung bình ($\bar{d}$)

Độ lệch tuyệt đối trung bình là số trung bình cộng của các độ lệch tuyệt đối giữa các lượng biến và số trung bình cộng của các lượng biến đó.

Công thức tính:

- Khi lượng biến xuất hiện một lần:

$$\bar{d} = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n}$$

- Khi lượng biến xuất hiện nhiều lần:

$$\bar{d} = \frac{\sum_{i=1}^m |x_i - \bar{x}| f_i}{\sum_{i=1}^m f_i}$$

Trong đó: $\bar{d}$ - Độ lệch tuyệt đối trung bình;

x_i - Các lượng biến của đơn vị thứ i(hoặc tổ i);

$\bar{x}$ - Trung bình cộng của các lượng biến x_i .

f_i - Tần số của tổ thứ i.

Khi độ lệch tuyệt đối trung bình càng nhỏ thì tính chất đồng đều của tổng thể càng lớn, tính chất đại biểu của số trung bình số học càng cao và ngược lại. Độ lệch tuyệt đối trung bình có ưu điểm hơn khoảng biến thiên vì số đo này lưu ý đến tất cả các lượng biến trong dãy số.

Ví dụ: Theo số liệu trong ví dụ của điểm a nêu trên ta có:

$$\bar{d}_A = \frac{|(40 - 60) + (50 - 60) + (70 - 60) + (80 - 60)|}{5} = 12$$

$$\bar{d}_B = \frac{|(58 - 60) + (59 - 60) + (61 - 60) + (62 - 60)|}{5} = 1,2$$

Qua kết quả tính toán, cũng có thể kết luận như điểm a nêu trên.

3.4.2.3. Phương sai

Phương sai là số trung bình cộng của bình phương độ lệch giữa các lượng biến x_i với số trung bình cộng của các lượng biến đó ($\bar{x}$).

Công thức tính:

$$\delta^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} \quad \text{hoặc} \quad \delta^2 = \frac{\sum_{i=1}^m (x_i - \bar{x})^2 f_i}{\sum_{i=1}^m f_i}$$

Trong đó: δ^2 - Là phương sai;

Phương sai càng nhỏ thì tính chất đồng đều của tổng thể càng cao, tính chất đại biểu của số trung bình số học càng cao và ngược lại.

Theo số liệu trong ví dụ của điểm a nêu trên ta có thể tính:

$$\bar{\delta}_A = \frac{(40 - 60)^2 + (50 - 60)^2 + (70 - 60)^2 + (80 - 60)^2}{5} = 200$$

$$\bar{\delta}_B = \frac{(58 - 60)^2 + (59 - 60)^2 + (61 - 60)^2 + (62 - 60)^2}{5} = 2$$

Qua công thức tính phương sai cho thấy sự khác nhau về dấu giữa các độ lệch đã được khắc phục bằng cách tính bình phương. Song cũng chính vì thế mà trị số bị khuếch đại, đơn vị tính không phù hợp với thực tế. Để khắc phục nhược điểm này, người ta đưa ra chỉ tiêu độ lệch tiêu chuẩn.

3.4.2.4. Độ lệch tiêu chuẩn (δ)

Độ lệch tiêu chuẩn là căn bậc 2 của phương sai.

$$\text{Công thức tính: } \delta = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} \quad \text{hoặc} \quad \delta = \sqrt{\frac{\sum_{i=1}^m (x_i - \bar{x})^2 \times f_i}{\sum_{i=1}^m f_i}}$$

Chỉ tiêu độ lệch tiêu chuẩn là chỉ tiêu hoàn thiện nhất và thường dùng nhất trong nghiên cứu thống kê để đánh giá độ biến thiên của tiêu thức. Tuy nhiên việc tính toán là khá phức tạp.

Vẫn lấy số liệu trong ví dụ của điểm a nêu trên ta có:

$$\delta_A = \sqrt{200} = 14,4$$

$$\delta_B = \sqrt{2} = 1,4142$$

3.4.2.5. Hệ số biến thiên (V)

Hệ số biến thiên là số tương đối so sánh giữa độ lệch tuyệt đối trung bình ($\bar{d}$) hoặc độ lệch tiêu chuẩn (δ) với số trung bình số học của các lượng biến đó.

$$V = \frac{\bar{d}}{\bar{x}} 100 \quad \text{hoặc} \quad V = \frac{\delta}{\bar{x}} 100$$

Chỉ tiêu hệ số biến thiên cho phép so sánh hai tổng thể không cùng quy mô, hoặc so sánh giữa hai chỉ tiêu thống kê khác nhau.

Theo số liệu trong ví dụ của điểm a nêu trên ta có:

Đối với bưu cục A:

$$V_1^A = (12 : 60) \times 100 = 20\%$$

$$V_2^A = (14,14 : 60) \times 100 = 23,56\%$$

Đối với bưu cục B:

$$V_1^B = (1,2 : 60) \times 100 = 2\%$$

$$V_2^B = (1,4142 : 60) \times 100 = 2,356\%$$

3.4.3. Các phương pháp tính phương sai

3.4.3.1. Các tính chất toán học của phương sai

Tính chất 1: Phương sai của một hằng số bằng 0

$$\delta_c^2 = 0$$

Tính chất 2: Nếu tất cả các lượng biến x_i được cộng (hoặc trừ) đi một số x_0 bất kỳ thì phương sai tính ra không đổi:

$$\delta_{(x_i + x_0)}^2 = \delta_x^2$$

Trong đó: $\delta_{(x_i + x_0)}^2$ - Phương sai của các lượng biến $x_i \pm x_0$

δ_x^2 - Phương sai của các lượng biến x_i

Tính chất 3: Nếu tất cả các lượng biến x_i chia cho cùng một số d thì phương sai tính ra sẽ giảm đi d^2 lần.

$$\delta_x^2 = d^2 \times \delta_{x/d}^2$$

Trong đó: $\delta_{x/d}^2$ - Phương sai các lượng biến x_i đã chia cho cùng một số d.

Tính chất 4: Độ lệch bình phương bình quân giữa các lượng biến x_i với số bình quân cộng $\bar{x}$ bao giờ cũng nhỏ hơn độ lệch bình phương bình quân giữa các lượng biến x_i với một trị số A bất kỳ khác số bình quân cộng $\bar{x}$. Nghĩa là phương sai có tính chất cực tiểu:

$$\delta_A^2 = \delta_x^2 + (\bar{x} - A)^2$$

Trong đó: δ_A^2 - Độ lệch bình phương bình quân giữa các lượng biến x_i với trị số A

3.4.3.2. Các phương pháp tính phương sai

Như trên đã nêu công thức tính phương sai theo định nghĩa nhưng khi dãy số lượng biến có các giá trị lớn thì tính toán sẽ không thích hợp. Do vậy có thể tính phương sai bằng các phương pháp khác.

a) Tính phương sai theo công thức biến đổi

Cách tính này được suy ra từ công thức phần định nghĩa.

$$\delta_x^2 = \overline{x^2} - \bar{x}^2$$

Trong đó: $\overline{x^2}$ - Bình quân cộng của các bình phương x_i^2

$$\overline{x^2} = \frac{\sum x_i^2 f_i}{\sum f_i}$$

$\bar{x}^2$ - Bình phương của số bình quân cộng x_i

$$\bar{x}^2 = \left[\frac{\sum x_i f_i}{\sum f_i} \right]$$

Ví dụ: Tính phương sai theo số liệu bảng 3.2. Để tiện so sánh sẽ tính theo định nghĩa và tính theo biến đổi. Lập bảng tính toán sau:

Bảng 3.4

% thực hiện kế hoạch	Trị số giữa x_i	Số đơn vị f_i	Tính theo định nghĩa		Tính theo biến đổi	
			$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 \cdot f_i$	x_i^2	$x_i^2 f_i$
75 - 85	80	6	1095,61	6.573,66	6.400	38.400
85 - 95	90	8	533,61	4.268,88	8.100	64.800
95 - 105	100	20	171,61	3.432,20	10.000	200.000
105 - 115	110	24	9,61	230,64	12.100	290.400
115 - 125	120	18	47,61	856,98	14.400	259.200
125 - 135	130	11	285,61	341,71	16.900	185.900
135 - 145	140	7	723,61	3.065,27	19.600	137.200
145 - 155	150	6	1.361,61	8.169,66	22.500	136.000
Cộng	-	100	-	31.739,00	-	1.310.900

Để tính bảng trên, ta đã có $n = \sum f_i = 100$ và $\bar{x} = 113,1$

Vậy phương sai tính theo:

- Công thức định nghĩa:

$$\delta_x^2 = \frac{31.739}{100} = 317,39$$

- Công thức biến đổi:

$$\delta_x^2 = \frac{31.739}{100} - (113,1)^2 = 317,39$$

Cả hai phương pháp tính đều cho kết quả như nhau, nhưng rõ ràng tính theo công thức biến đổi ngắn gọn hơn.

b) Tính phương sai bằng đổi biến

Khi tài liệu phân tổ phức tạp, việc tính phương sai bằng công thức định nghĩa cũng như công thức biến đổi vẫn dài dòng và tốn kém thời gian. Để khắc phục có thể tính bằng phương pháp đổi biến. Phương pháp này dựa trên cơ sở một số tính chất toán học của phương sai.

Quá trình tính toán phương sai bằng phương pháp đổi biến tiến hành như sau:

- Tạo biến mới:

$$x_i' = \frac{x_i - x_0}{d}$$

Trong đó: x_0 và d là các hằng số tự chọn $\neq 0$

+ Tính phương sai theo biến mới (bằng công thức định nghĩa hay công thức biến đổi):

$$\delta_{x'}^2 = \frac{\sum (x_i' - \bar{x}')^2 \times f_i}{\sum f_i}$$

$$\delta_{x'}^2 = \overline{x'^2} - \bar{x'}^2$$

+ Tính phương sai theo biến cũ:

$$\delta_x^2 = \frac{1}{d^2} \times \delta_x^2 \text{ do đó } \delta_x^2 = d^2 \times \delta_x^2$$

Ví dụ: Vẫn theo tài liệu phân tổ ở phần trên, tính phương sai bằng phương pháp đổi biến.

- Tạo biến mới bằng cách chọn $x_0 = 110$; $d = 10$. Vậy

$$x_i' = \frac{x_i - 110}{10}$$

Lập bảng tính toán:

Bảng 3.5

% thực hiện kế hoạch	Trị số giữa x_i	Số đơn vị f_i	$x_i' = \frac{x_i - x_0}{d}$	$x_i'^2$	$x_i'^2 \times f_i$
75 - 85	80	6	-3	9	54
85 - 95	90	8	-2	4	32
95 - 105	100	20	-1	1	20
105 - 115	110	24	0	0	0
115 - 125	120	18	1	1	18
125 - 135	130	11	2	4	44
135 - 145	140	7	3	9	63
145 - 155	150	6	4	16	96
Cộng	-	100	-	-	327

- Tính phương sai theo biến mới bằng công thức biến đổi:

Ta đã có $\overline{x'} = 0,31$, do đó

$$\delta_{x'}^2 = \overline{x'^2} - \overline{x'}^2 = \frac{327}{100} - (0,31)^2 = 3,1739$$

- Tính phương sai theo biến cũ

$$\delta_x^2 = d^2 \times \delta_x'^2 = 10^2 \times 3,1739 = 317,39$$

) Phương sai tiêu thức chất lượng

Trong số các tiêu thức chất lượng thường gặp loại tiêu thức là biểu hiện cụ thể của nó chỉ có 02 trường hợp: có hoặc không, nghĩa là ở một số đơn vị tổng thể có biểu hiện đặc tính quan tâm, còn một số đơn vị còn lại thì không biểu hiện.

Để xây dựng công thức tính số bình quân và phương sai cho loại tiêu thức này ta quy ước một số ký hiệu:

$x_1 = 1$ khi đơn vị quan sát có biểu hiện đặc tính quan tâm.

$x_2 = 0$ khi đơn vị quan sát không biểu hiện

p - Tỷ trọng của bộ phận có biểu hiện

q - Tỷ trọng của bộ phận không có biểu hiện.

p và q là những số tương đối nhỏ hơn 1, giữa chúng có mối quan hệ $p + q = 1$. Xác định số bình quân và phương sai căn cứ vào công thức định nghĩa.

- Số bình quân cộng

$$\bar{x} = \frac{\sum x_i \times f_i}{\sum f_i} = \frac{x_1 \times p + x_2 \times q}{p + q} = \frac{1 \times p + 0 \times q}{1} = p$$

- Phương sai

$$\delta_x^2 = \frac{\sum (x_i - \bar{x})^2 \times f_i}{\sum f_i} = \frac{(1 - p)^2 \times p + (0 - p)^2}{p + q} = \frac{q^2 \times p + p^2 \times q}{p + q}$$

$$\frac{p \cdot q \times (p + q)}{p + q} = p \times q$$

3.5. PHÂN PHỐI TRONG THỐNG KÊ

3.5.1. Một số phân phối lý thuyết

3.5.1.1. Phân phối nhị thức

a) Công thức

Giả sử thực hiện m phép thử độc lập. Trong mỗi phép thử biến cố A xuất hiện với xác suất $P(A) = p$. Khi đó xác suất để biến cố A xuất hiện đúng x lần trong m phép thử được xác định bằng công thức:

$$P(x) = C_m^x p^x (1-p)^{m-x}$$

Trong đó: x - Các đại lượng ngẫu nhiên có thể nhận các giá trị $0, 1, 2, 3, \dots, m$.

$(1-p)$ - Xác suất không xuất hiện biến cố A trong mỗi phép thử $1 - p = q$.

C_m^x - Các hệ số của phân phối nhị thức (tổ hợp chập x của tập hợp gồm m phần tử). Nó được xác định:

$$C_m^x = \frac{m!}{x!(m-x)!}$$

$$\text{Nhu vậy: } P(x) = \frac{m!}{x!(m-x)!} p^x q^{m-x}$$

b) Tính chất (không chứng minh)

- Tổng các xác suất tính từ mọi giá trị có thể của x bằng 1

$$\sum_{x=0}^m P(x) = 1$$

$$P(x) = \frac{\lambda^x \cdot e^{-\lambda}}{x!}$$

Trong đó: $e = 2,718...$ Cơ số của lô-ga-rít

$x = 0, 1, 2, \dots$ Đại lượng ngẫu nhiên rời rạc

b) Các tham số

- Số bình quân: $\bar{x} = \lambda = mp$
- Phương sai: $\sigma_x^2 = \lambda$
- Độ lệch chuẩn: $\sigma_x = \sqrt{\lambda}$

c) Tính chất

Tổng các xác suất tính từ mọi giá trị có thể của x bằng 1

$$\sum_{x=0}^m (P(x)) = 1$$

Việc xác định các giá trị $P(x)$ đã được ghi sẵn.

3.5.1.3. Phân phối chuẩn

Phân phối chuẩn sử dụng để xét các đại lượng ngẫu nhiên liên tục.

a) Phương trình đường cong phân phối chuẩn

Phân phối chuẩn là một phân phối liên tục. Trên đồ thị các tung độ của đường cong được xác định:

$$f(x) = \frac{1}{\sigma_x \sqrt{2\pi}} e^{-\frac{(x-\bar{x})^2}{2\sigma_x^2}}$$

Trong đó: $\pi = 3,14159$; $e = 2,718$;

x - Đại lượng ngẫu nhiên liên tục có thể lấy mọi giá trị từ $-\infty$ đến $+\infty$

$\bar{x}$ - Số bình quân; σ_x - Độ lệch chuẩn.

Đường cong $f(x)$ có dạng hình chuông đối xứng. Tại giá trị $x = \bar{x}$ tung độ đạt giá trị cực đại.

$$f(x) = f(\bar{x}) = \frac{1}{\sigma_x \sqrt{2\pi}} e^0 = \frac{1}{\sigma_x \sqrt{2\pi}}$$

Đỉnh của hình chuông là P ứng với tung độ cực đại. Ở hai bên của P tung độ giảm dần ; lúc đầu giảm ít, sau giảm nhiều, đến 02 điểm đối xứng ứng với hoành độ $\bar{x} - \sigma_x$ và $\bar{x} + \sigma_x$ là 02 điểm uốn đường cong. Sau đó tung độ giảm chậm và cuối cùng bên trái cũng như bên phải đường cong tiệm cận với trục hoành.

b) Các tham số đặc trưng

- Số bình quân: $\bar{x} = \int_{-\infty}^{+\infty} xf(x)dx$

- Phương sai: $\sigma_x^2 = \int_{-\infty}^{+\infty} (x - \bar{x})^2 f(x)dx$

- Độ lệch chuẩn: $\sigma_x = \sqrt{\sigma_x^2}$

Số bình quân xác định vị trí đường cong trên trục x , còn độ lệch chuẩn xác định hình dáng nhọn hay dẹt của đường cong (Độ lệch chuẩn bé thì đường cong nhọn, độ lệch chuẩn lớn thì đường cong dẹt).

Vì đường cong đối xứng qua $\bar{x}$ nên số bình quân vừa là số mốt, vừa là số trung vị $\bar{x} = M_0 = M_c$

Hàm $f(x)$ là hàm mật độ xác suất nên có thể chứng minh được:

$$\int_{-\infty}^{+\infty} f(x) dx = 1$$

c) Đường cong chuẩn thu gọn

Nếu trong công thức đặt:

$$t = \frac{x - \bar{x}}{\sigma_x}$$

Khi đó số bình quân theo biến mới (t) sẽ là:

$$\bar{t} = \frac{\bar{x} - \bar{x}}{\sigma_x} = 0$$

Các phương sai theo biến t:

$$\sigma_t^2 = \sigma_{\frac{x - \bar{x}}{\sigma_x}}^2 = \frac{1}{\sigma_x^2} \sigma_{(x)}^2 = 1$$

Như vậy từ cách biến đổi công thức có thể viết thành dạng:

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

Hàm $\varphi(t)$ được gọi là phương trình đường cong chuẩn thu gọn, còn t được gọi là biến chuẩn hóa. Ở đây t được phân phối theo luật chuẩn với số bình quân $\bar{t} = 0$ và độ lệch chuẩn $\sigma_t = 1$.

3.5.2. So sánh phân phối thực nghiệm với phân phối lý thuyết

3.5.2.1. So sánh bằng tiêu chuẩn λ^2

Khi so sánh phân phối thực nghiệm với phân phối lý thuyết bằng tiêu chuẩn λ^2 , ta tiến hành theo các bước sau:

Bước 1: Vẽ đồ thị tần số thực nghiệm qua đó phán đoán dạng phân phối lý thuyết

Do có nhiều dạng phân phối lý thuyết, người ta thường chỉ chọn ra một dạng để đem so sánh, vì vậy để đỡ tốn thời gian phải chọn nhiều lần, ngay từ đầu cần có sự phán đoán tương đối chính xác dạng phân phối lý thuyết. Mỗi dạng phân phối lý thuyết có một đồ thị biểu diễn nhất định. Thông qua đồ thị tần số thực nghiệm sẽ có sự liên hệ và căn cứ hợp lý để phán đoán. Tuy nhiên điều phán đoán được vẫn không thể thay thế được kết luận cuối cùng.

Bước 2: Tính các tham số cần thiết của phân phối thực nghiệm

- Xác suất p, q (đối với phân phối lý thuyết và phân phối nhị thức)
- Số bình quân
- Độ lệch chuẩn
- Xác suất P(x)
- Tần số lý thuyết f_i

Cách xác định các tham số này đã được trình bày cụ thể trong phần trên.

Bước 3: Tính giá trị tiêu chuẩn λ^2 và so sánh kết luận

Giá trị tiêu chuẩn λ^2 được tính theo công thức:

$$\lambda^2 = \sum_{i=1}^k \frac{(f_i - f_i')^2}{f_i'}$$

Trong đó: f_i - Tần số thực nghiệm

f_i' - Tần số lý thuyết

$i = 1, 2, \dots, k$ - Thứ tự các tổ.

Chú ý: Khi áp dụng công thức cần thỏa mãn điều kiện trong tất cả các tổ tần số lý thuyết cũng như tần số thực nghiệm đều phải lớn hơn hoặc bằng 5. Nếu chưa thỏa mãn cần phải tiến hành ghép tổ.

Sau khi tính được λ^2 đem so sánh với giá trị $\lambda_{\text{bảng}}^2$ (tra được từ bảng tính sẵn với độ tự do T và hệ số sai lầm cho phép P_s)

- Nếu $\lambda^2 < \lambda_{\text{bảng}}^2$ thì kết luận phân phối thực nghiệm phù hợp với phân phối lý thuyết với hệ số sai lầm P_s .

- Nếu $\lambda^2 \geq \lambda_{\text{bảng}}^2$ thì không có cơ sở kết luận phù hợp của 02 phân phối. Lúc đó tạm thời coi phân phối thực nghiệm không phù hợp với phân phối lý thuyết với hệ số sai lầm P_s .

3.5.2.2. So sánh bằng tiêu chuẩn Romanosky

Ở đây các bước tiến hành cũng giống như khi sử dụng tiêu chuẩn λ^2 , nhưng để đánh giá kết quả và rút ra kết luận cuối cùng người ta không sử dụng $\lambda_{\text{bảng}}^2$ mà dùng chỉ tiêu R.

$$R = \frac{|\lambda^2 - T|}{\sqrt{2T}}$$

Trong đó: λ^2 - Tính theo công thức ở bước 3, mục 3.5.2.1

T - Độ tự do.

Căn cứ vào giá trị của R để kết luận:

- Nếu $R < 3$ thì phân phối thực nghiệm phù hợp với phân phối lý thuyết
- Nếu $R \geq 3$ thì phân phối thực nghiệm không phù hợp với phân phối lý thuyết.

3.5.2.3. So sánh bằng tiêu chuẩn Kolmogorop

Việc so sánh cũng được tiến hành theo ba bước, trong đó bước 1 và bước 2 giống như khi so sánh bằng tiêu chuẩn χ^2 .

Bước 3: Tính giá trị tiêu chuẩn và so sánh. Tiêu chuẩn Kolmogorop D được xác định bằng công thức:

$$D = \frac{1}{n} \max |F_i - F_i'|$$

Trong đó: F_i - Tần số thực nghiệm cộng dồn

F_i' - Tần số lý thuyết cộng dồn

n - Tổng các tần số $n = \sum f_i = \sum f_i'$

Giá trị của D tính theo công thức ở bước 3, mục 3.5.2.3 được so sánh với $D_{\text{bảng}}$

- Nếu $D < D_{\text{bảng}}$ thì kết luận với hệ số sai lầm P , phân phối thực nghiệm phù hợp với phân phối lý thuyết.
- Nếu $D \geq D_{\text{bảng}}$ thì phân phối thực nghiệm không phù hợp với phân phối lý thuyết.

Khi sử dụng tiêu chuẩn Kolmogorop không cần phải ghép các tổ có tần số nhỏ hơn 5.

3.5.3. Ứng dụng các trường hợp

3.5.3.1. So sánh phân phối thực nghiệm với phân phối nhị thức (sử dụng tiêu chuẩn χ^2)

Ví dụ: Số dịch vụ EMS có trong lượng dịch vụ khách hàng đến yêu cầu phục vụ tại một bưu cục là một đại lượng ngẫu nhiên x. Giả sử các giá trị có thể của nó là 0, 1, 2, 3, 4, 5, 6. Quan sát 218 khách hàng đến yêu cầu ta có số liệu sau:

Bảng 3.6

Số dịch vụ EMS	Số khách hàng đến
0	18
1	48
2	62
3	40
4	27
5	15
6	8
Cộng	218

Số dịch vụ EMS đến có tuân theo phân phối nhị thức không?

Quá trình so sánh được tiến hành theo các bước:

Bước 1: Vẽ đồ thị tần số và phán đoán dạng phân phối lý thuyết. Việc vẽ đồ thị nhằm mục đích phán đoán dạng phân phối lý thuyết thích hợp. Trong ví dụ này phân phối lý thuyết đã được biết trước nên không cần vẽ đồ thị.

Bước 2: Tính các tham số của phân phối thực nghiệm

- Số dịch vụ EMS đến bình quân

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{523}{218} = 2,399$$

Bảng 3.7

x_i	f_i	$x_i f_i$	$P(x) = C_m^x \times P^x \cdot q^{m-x}$	$f_i' = n.P(x)$	$f_i - f_i'$	$(f_i - f_i')^2$	$\frac{(f_i - f_i')^2}{f_i'}$
0	18	0	0,046656	10	8	64	6,400
1	48	48	0,186624	41	7	49	1,195
2	62	124	0,311040	68	- 6	36	0,529
3	40	120	0,276480	60	- 20	400	0,667
4	27	108	0,138240	30	- 3	9	0,300
5	15	75	0,036864	8	14	196	21,778
6	8	48	0,004096	1			
Cộng	218	523	1,000000	218	-	-	$\lambda^2 = 36,869$

- Xác suất p là xác suất để xuất hiện dịch vụ EMS trong mỗi khách hàng đến yêu cầu.

$$\bar{x} = mp \Rightarrow p = \frac{\bar{x}}{m} = \frac{2,399}{6} = 0,4$$

$$q = 1 - p = 1 - 0,4 = 0,6$$

- Xác suất P(x) tức là xác suất để trong mỗi lần khách hàng đến có 0, 1, 2, 3, 4, 5, 6 dịch vụ EMS. Nó được tính theo công thức ở mục 3.5.1.1 (a).

$$P(x) = C_6^x \cdot 0,4^x \cdot 0,6^{6-x}$$

Giá trị của $P(x)$ được tính ở cột trong bảng.

- Tần số lý thuyết f_i' .

$$f_i' = n \cdot P(x) = 218 \cdot P(x)$$

Giá trị của nó được ghi ở cột trong bảng.

Bước 3: Tính giá trị λ^2 và so sánh

Để tính λ^2 trước hết cần tính hiệu $(f_i - f_i')$; $(f_i - f_i')^2$; và $(f_i - f_i')^2 / f_i'$. Tổng các giá trị cột chính là λ^2 . Kết quả $\lambda^2 = 36,869$. Khi tính cần chú ý tần số lý thuyết tổ cuối cùng bằng 1 nhỏ hơn 5 nên phải ghép với tổ đứng trước nó.

Tra bảng giá trị λ_{bang}^2 với hệ số sai lầm $P_s = 0,05$ và độ tự do:

$$T = k - 2 = 7 - 2 = 5$$

$$\text{Có: } \lambda_{\text{bang}}^2 = 11,07$$

Như vậy λ^2 tính toán $= 36,869 > \lambda_{\text{bang}}^2$ do đó kết luận phân phối thực nghiệm không phù hợp với phân phối lý thuyết.

Trường hợp sử dụng tiêu chuẩn so sánh khác, chẳng hạn Romanosky ta cũng có kết luận tương tự.

$$R = \frac{|\lambda^2 - T|}{\sqrt{2T}} = \frac{36,869 - 4}{\sqrt{2 \cdot 4}} = 11,6 > 3$$

3.5.3.2. So sánh phân phối thực nghiệm với phân phối Poisson

Ví dụ: Khảo sát phân phối thực nghiệm thời gian đàm thoại của 100 khách hàng dưới đây tuân theo phân phối lý thuyết nào?

Bảng 3.8

Thời gian đàm thoại (phút)	Số khách hàng quan sát	Thời gian đàm thoại (phút)	Số khách hàng quan sát
0	2	6	10
1	10	7	6
2	14	8	3
3	20	9	2
4	18	10	1
5	13	11	1
Cộng			100

Bước 1: Vẽ đồ thị tần số thực nghiệm và phán đoán dạng phân phối

Qua đồ thị cho thấy các tần số thực nghiệm được phân phối tương đối giống dạng phân phối Poisson. Vì vậy tiến hành so sánh với phân phối này.

Bước 2: Tính các tham số của phân phối thực nghiệm

Thời gian đàm thoại bình quân:

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{400}{100} = 4 \text{ phút}$$

Đây là tham số đặc trưng của phân phối Poisson ($\bar{x} = \lambda$)

Bảng 3.9

Thời gian đàm thoại x_i	Số khách hàng f_i	$x_i f_i$	$P(x)$	f_i'	$f_i - f_i'$	$(f_i - f_i')^2$	$\frac{(f_i - f_i')^2}{f_i'}$
0	2	0	0,0210	2	3	9	1,000
1	10	10	0,0724	7			

Thời gian đàm thoại x_i	Số khách hàng f_i	$x_i f_i$	$P(x)$	f_i'	$f_i - f_i'$	$(f_i - f_i')^2$	$\frac{(f_i - f_i')^2}{f_i'}$
2	14	28	0,1465	15	- 1	1	0,066
3	20	60	0,1954	20	0	0	0
4	18	72	0,1954	20	- 2	4	0,200
5	13	65	0,1563	16	- 3	9	0,562
6	10	60	0,1042	10	0	0	0
7	6	42	0,0595	6	0	0	0
8	3	24	0,0298	3	3	9	0,900
9	2	18	0,0132	1			
10	1	10	0,0053				
11	1	11	0,0019				
Cộng	100	400	1,000	100	-	-	$\chi^2 = 2,728$

- Xác suất $P(x)$ tức là xác suất để với mỗi khách hàng đàm thoại nhận các giá trị $x = 0, 1, 2, \dots, 10, 11$.

Giá trị $P(x)$ được xác định theo công thức

$$P(x) = \frac{4^x \cdot e^{-4}}{x!}$$

Kết quả tính toán ghi trong bảng:

- Tần số lý thuyết f_i' :

$$f_i' = n \cdot P(x) = 100 \cdot P(x)$$

Kết quả tính ghi vào bảng.

Bước 3: Tính giá trị tiêu chuẩn và so sánh

Để tính giá trị tiêu chuẩn λ^2 ta cần tính $(f_i - f_i')^2$; $(f_i - f_i')^2$ và $(f_i - f_i')^2 / f_i'$. Kết quả ghi trong bảng. Ô cuối cùng cho ta giá trị $\lambda^2 = 2,728$.

Chú ý: Theo quy định tần số phải lớn hơn hoặc bằng 5 do đó ta phải ghép tổ 1 và 2; các tổ 8, 9, 10, 11 với độ tự do $T = k - 2 = 7 - 2 = 5$ và $P_s = 0,05$ ta tìm trong bảng giá trị $\lambda_{\text{bảng}}^2 = 11,07$. Như vậy:

$$\lambda_{\text{tính toán}}^2 = 2,728 < \lambda_{\text{bảng}}^2 = 11,07$$

Kết luận: Khảo sát 100 khách hàng cho thấy thời gian đàm thoại tuân theo phân phối Poisson.

3.5.3.3. So sánh phân phối thực nghiệm với phân phối chuẩn

Ví dụ: Quan sát 50 bức điện báo ta có số tiếng trong mỗi bức như sau:

Bảng 3.10

Số tiếng x_i	Bức điện báo f_i	Trị số giữa x_i
19 - 21	2	20
21 - 23	5	22
23 - 25	6	24
25 - 27	10	26
27 - 29	11	28
29 - 31	9	30
31 - 33	6	32
33 - 35	1	34
Cộng	50	

Yêu cầu khảo sát phân phối thực nghiệm này có tuân theo phân phối chuẩn hay không?

* Trường hợp so sánh bằng tiêu chuẩn λ^2 :

Bước 1: Vẽ đồ thị và phán đoán dạng phân phối lý thuyết.

Trong trường hợp này phân phối lý thuyết (phân phối chuẩn) đã được chọn trước. Cho nên không cần vẽ đồ thị.

Bước 2: Tính các tham số của phân phối thực nghiệm

- Số tiếng bình quân:

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{1.358}{50} = 27,16$$

Bảng 3.11

Số tiếng (phút) x_i	Bức điện f_i	$x_i \cdot f_i$	$x_i^2 \cdot f_i$	t_i	$\phi(t_i)$	f_i'	$f_i - f_i'$	$(f_i - f_i')^2$	$\frac{(f_i - f_i')^2}{f_i'}$
20	2	40	800	- 2,11	0,0431	1	2	4	0,667
22	5	110	2420	- 1,52	0,1257	4			
24	6	144	3456	- 0,93	0,2589	8	- 2	4	0,500
26	10	260	6760	- 0,34	0,3765	11	- 1	1	0,090
28	11	308	8624	0,25	0,3867	12	- 1	1	0,833
30	9	270	8100	0,84	0,2803	8	1	1	0,125
32	6	192	6144	1,43	0,1435	4	1	1	0,167
34	1	34	1156	1,85	0,0721	2			
Cộng	50	1358	37460	-	-	50	-	-	$\lambda^2 = 2,382$

- Phương sai: $\sigma_x^2 = \overline{x^2} - \bar{x}^2 = \frac{37.460}{50} - 27,16^2 = 11,53$

- Độ lệch chuẩn: $\sigma_x = \sqrt{11,53} = 3,396$

- Xây dựng hàm $\varphi(t_i)$:

$$+ \text{Tính biến mới: } t_i = \frac{x_i - \bar{x}}{\sigma_x} = \frac{x_i - 27,16}{3,396}$$

Kết quả tính ghi trong bảng.

+ Tra bảng tìm các giá trị $\varphi(t_i)$.

- Tần số lý thuyết n_i' :

$$f_i' = \frac{n \times h}{\sigma_x} \varphi(t_i) = \frac{50 \times 2}{3,396} \varphi(t_i) = 29,446 \times \varphi(t_i)$$

Các giá trị của f_i' được ghi trong bảng.

Bước 3: Tính giá trị tiêu chuẩn và so sánh

Việc tính tiêu chuẩn λ^2 tương tự như các ví dụ trước. Kết quả tính ta có $\lambda^2 = 2,382$.

Chú ý: Trong ví dụ này có 04 tổ tần số nhỏ hơn 5 nên phải ghép.

Tra bảng tìm giá trị λ_{bang}^2 với sai số $P_s = 0,05$ và độ tự do:

$$T = k - 3 = 6 - 3 = 3 \text{ có } \lambda_{\text{bang}}^2 = 7,82$$

Như vậy: $\lambda_{\text{tinh toan}}^2 = 2,382 < \lambda_{\text{bang}}^2 = 7,82$. Chúng ta phân phối thực nghiệm số tiếng của 50 bưu điện báo tuân theo phân phối chuẩn.

* Trường hợp so sánh bằng tiêu chuẩn Kolmogorop (D):

Bước 1 và 2: Như tiêu chuẩn λ^2 .

Bước 3: Tính giá trị tiêu chuẩn và so sánh.

Để tiện cho việc theo dõi lập bảng tính toán sau:

Bảng 3.12

x_i	f_i	f_i'	F_i	F_i'	$F_i - F_i'$
20	2	1	2	1	1
22	5	4	7	5	2
24	6	8	13	13	0
26	10	11	23	24	- 1
28	11	12	34	36	- 2
30	9	8	43	44	- 1
32	6	4	49	48	1
34	1	2	50	50	0
Cộng	50	50	-	-	-

Qua bảng ta tính:

$$D = \frac{1}{n} \max |F_i - F_i'|$$

$$= \frac{1}{50} \times 2 = \frac{2}{50} = 0,04$$

Tra bảng giá trị của $D_{\text{bảng}}$ với $P_s = 0,05$, ta có:

$$D_{\text{bảng}} = \frac{1,36}{\sqrt{50}} \quad (\text{vì } n = 50 > 40)$$

$$= \frac{1}{50} \times 2 = \frac{2}{50} = 0,04$$

Vậy $D_{\text{tính toán}} = 0,04 < D_{\text{bảng}} = 0,192$

Kết luận vẫn như trên, nghĩa là quan sát 50 bức điện báo cho thấy thời gian đàm thoại tuân theo phân phối chuẩn.

Chương 4

ĐIỀU TRA CHỌN MẪU

4.1. KHÁI NIỆM, ƯU NHƯỢC ĐIỂM VÀ PHẠM VI SỬ DỤNG ĐIỀU TRA CHỌN MẪU

4.1.1. Khái niệm điều tra chọn mẫu

Điều tra chọn mẫu là một loại điều tra không toàn bộ trong đó người ta chỉ chọn ra một số đơn vị trong toàn bộ các đơn vị của hiện tượng nghiên cứu để điều tra thực tế, rồi dùng các kết quả tính toán để suy rộng thành các đặc điểm của toàn bộ tổng thể nghiên cứu.

Tại sao chỉ điều tra một số đơn vị tổng thể mà suy ra kết quả cho cả tổng thể. Vì quy luật số lớn đã chỉ ra rằng nếu nghiên cứu một số tương đối lớn hiện tượng thì những biểu hiện ngẫu nhiên, những đặc thù của hiện tượng riêng lẻ sẽ bù trừ và triệt tiêu cho nhau, tính quy luật sẽ được biểu hiện rõ. Hơn nữa lý thuyết xác suất cũng chứng minh rằng sự sai khác giữa số bình quân của một số rất lớn đại lượng ngẫu nhiên với kỳ vọng toán của nó là một đại lượng nhỏ tùy ý.

Khi chọn đơn vị để điều tra chọn mẫu người ta có thể chọn theo cách ngẫu nhiên, nghĩa là các đơn vị tổng thể đều có khả năng được chọn như nhau không phụ thuộc vào ý kiến chủ quan người chọn mẫu, hoặc không ngẫu nhiên (bàn bạc, phân tích tập thể để lựa chọn ra những đơn vị đại biểu).

4.1.2. Ưu điểm của điều tra chọn mẫu

- Điều tra chọn mẫu thường nhanh hơn rất nhiều so với điều tra toàn bộ, vì điều tra ít đơn vị nên công tác chuẩn bị nhanh gọn, số lượng tài liệu ghi chép ít, thời gian điều tra ghi chép, tổng hợp và phân tích sẽ được rút ngắn. Do đó điều tra chọn mẫu mang tính kịp thời cao.

- Do điều tra ít đơn vị, số nhân viên điều tra và mọi chi phí điều tra giảm, cho nên điều tra chọn mẫu sẽ tiết kiệm sức người, vật tư, tiền của.

- Do điều tra ít đơn vị nên có thể mở rộng được nội dung điều tra, đi sâu nghiên cứu nhiều mặt của hiện tượng.

- Tài liệu thu thập được trong điều tra chọn mẫu sẽ có độ chính xác cao, bởi vì số nhân viên điều tra ít nên có thể chọn được những người có trình độ chuyên môn cao và nhiều kinh nghiệm, đồng thời việc kiểm tra số liệu có thể tiến hành tỷ mỉ và tập trung, do đó các sai sót do ghi chép sẽ giảm đi nhiều.

- Điều tra chọn mẫu không đòi hỏi một tổ chức lớn như điều tra toàn bộ, một tổ chức nhỏ cũng có thể tiến hành điều tra chọn mẫu.

Tóm lại điều tra chọn mẫu có rất nhiều ưu điểm so với điều tra toàn bộ, nhưng do điều tra chọn mẫu tiến hành với phạm vi nhỏ nên sẽ có sai số nhất định so với kết quả điều tra toàn bộ nên không thể dùng điều tra chọn mẫu để hoàn toàn thay thế cho điều tra toàn bộ.

4.1.3. Phạm vi ứng dụng điều tra chọn mẫu

- Dùng để thay thế điều tra toàn bộ khi đối tượng nghiên cứu cho phép vừa có thể điều tra toàn bộ, vừa có thể điều tra chọn

mẫu, thì tiến hành điều tra chọn mẫu để có kết quả nhanh và tiết kiệm hoặc dùng cho những trường hợp việc điều tra có liên quan tới phá hủy đơn vị điều tra.

- Kết hợp với điều tra toàn bộ để mở rộng nội dung điều tra và đánh giá kết quả điều tra toàn bộ.

- Dùng trong trường hợp muốn so sánh các hiện tượng với nhau mà chưa có tài liệu cụ thể, hoặc kiểm định giả thiết đặt ra.

- Dùng để tổng hợp nhanh tài liệu điều tra toàn bộ, có thông tin nhanh, phục vụ kịp thời cho công tác quản lý.

4.1.4. Phân loại điều tra chọn mẫu

4.1.4.1. Điều tra chọn mẫu ngẫu nhiên

Điều tra chọn mẫu ngẫu nhiên là phương pháp chọn mẫu trong đó các đơn vị tổng thể được chọn vào mẫu một cách hoàn toàn ngẫu nhiên, nghĩa là các đơn vị tổng thể đều có khả năng được chọn như nhau, không phụ thuộc vào ý muốn chủ quan người chọn mẫu.

4.1.4.2. Điều tra chọn mẫu phi ngẫu nhiên

Điều tra chọn mẫu phi ngẫu nhiên là sự lựa chọn các đơn vị vào mẫu điều tra dựa trên kinh nghiệm và hiểu biết của người nghiên cứu về tổng thể nghiên cứu, hoặc căn cứ vào những qui định nhất định khi lấy mẫu.

4.2. ĐIỀU TRA CHỌN MẪU NGẪU NHIÊN

4.2.1. Tổng thể chung và tổng thể mẫu

Tổng thể chung là tổng thể bao gồm toàn bộ các đơn vị thuộc đối tượng điều tra. Số đơn vị của tổng thể chung thường được ký hiệu là N và được biểu thị bằng dãy số lượng biến:

$$X_1, X_2, X_3, \dots, X_n, \dots, X_N$$

Tổng thể mẫu là tổng thể bao gồm một số đơn vị nhất định được chọn ra từ tổng thể chung để điều tra thực tế. Số đơn vị của tổng thể mẫu thường được ký hiệu là n . Tổng thể mẫu gồm n đơn vị tổng thể thường được biểu diễn dưới dạng bảng phân phối.

4.2.2. Chọn mẫu nhiều lần và chọn mẫu một lần

4.2.2.1. Chọn mẫu nhiều lần (chọn hoàn lại)

Chọn mẫu nhiều lần là khi mỗi đơn vị được chọn ra để đăng ký rồi lại trả về tổng thể chung. Như vậy số đơn vị tổng thể chung không thay đổi trong suốt quá trình lấy mẫu. Xác suất được chọn của mỗi đơn vị tổng thể là bằng nhau.

Gọi K là số khả năng thiết lập được tổng thể mẫu thì số khả năng đó trong chọn nhiều lần được xác định theo công thức sau:

$$K = N^n$$

4.2.2.2. Chọn mẫu một lần (chọn không hoàn lại)

Chọn mẫu một lần là khi mỗi đơn vị đã được chọn để đăng ký rồi sẽ được xếp riêng ra không trả lại về tổng thể chung nữa, do đó không có khả năng chọn lại, xác suất được chọn của mỗi đơn vị không bằng nhau.

Khi đó số khả năng thiết lập tổng thể mẫu được xác định theo công thức:

$$K = C_n^N = \frac{N!}{n!(N-n)!}$$

4.2.3. Chọn mẫu với xác suất đều và xác suất không đều

Chọn mẫu với xác suất đều là đảm bảo mỗi đơn vị của hiện tượng nghiên cứu đều có cơ hội được chọn vào mẫu như nhau. Tính bình đẳng còn thể hiện trong việc ước lượng kết quả vì kết quả thu được trên mẫu không phân biệt đơn vị nào. Phương pháp chọn mẫu không đều không lưu ý đến sự khác biệt giữa các đơn vị tổng thể và thường được sử dụng khi các đơn vị của tổng thể tương đối đồng đều nhau theo tiêu thức nghiên cứu, hoặc không biết trước được sự khác biệt giữa các đơn vị điều tra.

Chọn mẫu với xác suất không đều nghĩa là không cần đảm bảo khả năng được chọn vào mẫu của các đơn vị phải bằng nhau. Các đơn vị có thể được chọn vào mẫu theo xác suất tỷ lệ vai trò của từng đơn vị. Xác suất ấn định riêng cho mỗi đơn vị về khả năng được chọn vào mẫu là xác suất bao hàm.

Việc chọn mẫu với xác suất không đều có khó khăn, phức tạp hơn vì phải có số liệu tiên nghiệm về qui mô, kích cỡ của từng đơn vị.

4.2.4. Sai số trong điều tra chọn mẫu ngẫu nhiên

Sai số chọn mẫu là sự chênh lệch về trị số giữa các chỉ tiêu tính ra được trong điều tra chọn mẫu và các chỉ tiêu tương ứng của tổng thể chung.

4.2.4.1. Các loại sai số

Trong điều tra chọn mẫu ngẫu nhiên ngoài sai số do ghi chép, còn có sai số do tính chất đại biểu, sai số xảy ra do chỉ điều tra một số ít đơn vị tổng thể mà kết quả lại suy rộng cho cả tổng thể (sai số chọn mẫu).

a) Sai số chọn mẫu

Như đã nói, ta không biết được một cách chính xác các đặc trưng của tổng thể như trung bình hoặc tỷ lệ, mà chỉ có thể ước lượng chúng dựa trên thông tin thu thập từ mẫu. Các ước lượng đó, tất nhiên là không thể tránh khỏi sai sót, gọi là sai số chọn mẫu. Như vậy sai số chọn mẫu là sai số do sử dụng thông tin thu thập được chỉ trên một bộ phận tổng thể, hoặc là do mẫu không thể hiện, đại diện được cho cả tổng thể. Nói cách khác sai số chọn mẫu do hai yếu tố: yếu tố thiết kế, thực hiện và yếu tố ngẫu nhiên. Có thể hạn chế sai số bằng cách hết sức chú ý tới việc thiết kế, thực hiện chọn mẫu, song sai số chọn mẫu hầu như là tất nhiên, chỉ có thể giảm bớt bằng cách tăng qui mô của mẫu.

b) Sai số phi chọn mẫu

Sai số phi chọn mẫu là sai số không thuộc về phương pháp chọn mẫu được sử dụng. Các điều kiện để phát sinh sai số phi chọn mẫu có thể là:

- *Sai số do ghi chép*: Do đơn vị điều tra chưa hiểu đúng nội dung câu hỏi nên trả lời sai, do đo lường, do vô tình ghi chép sai hoặc cố ý ghi chép sai do mục đích nào đó.

Nếu là sai số ngẫu nhiên thì các sai lệch trong khi lấy số liệu có thể bù trừ nhau. Loại sai số này ít gây nguy hiểm.

Sai số hệ thống là sai số nguy hiểm, càng nhiều đơn vị điều tra thì sai số càng nhiều. Nó xảy ra do dụng cụ đo lường sai hay cố ý ghi sai.

Để có thể giảm bớt được sai số do ghi chép cần phải chuẩn bị tốt cho công tác điều tra, trình độ dụng cụ máy móc và ý thức người điều tra.

- Sai số do mẫu được lấy từ một tổng thể không thích hợp.

- Sai số do tỷ lệ không trả lời quá cao. Việc không nhận được các câu trả lời là một vấn đề quan trọng phải giải quyết của phương pháp chọn mẫu. Nó làm cho kết quả điều tra bị sai lệch do thông tin không đầy đủ, tổng thể trả lời có thể rất khác xa với tổng thể thực sự muốn điều tra.

4.2.4.2. Phương pháp tính sai số

a) Khi điều tra chọn mẫu dùng để suy rộng số trung bình của một tiêu thức nào đó thì sai số được tính như sau:

+ Khi chọn mẫu nhiều lần:

$$\mu_x = \sqrt{\frac{\delta^2}{n}}$$

Trong đó: μ_x - Sai số trung bình chọn mẫu.

δ^2 - Phương sai chung.

n - Số đơn vị tổng thể mẫu.

Do phương sai tổng thể chung không tính được nên phải dùng phương sai mẫu điều chỉnh để thay thế:

$$\mu_x = \sqrt{\frac{\delta_m^2}{n-1}}$$

+ Khi chọn mẫu một lần

$$\mu_x = \sqrt{\frac{\delta^2}{n} \left(1 - \frac{n}{N}\right)}$$

$$\text{Hay: } \mu_x = \sqrt{\frac{\delta_m^2}{n-1} \left(1 - \frac{n}{N}\right)}$$

Trong đó: δ_m^2 - Là phương sai mẫu điều chỉnh:

$$\delta_m^2 = \frac{n}{n-1} \delta_m^2$$

δ_m^2 - Phương sai mẫu.

N - Số đơn vị của tổng thể chung.

Ví dụ: Tính sai số chọn mẫu ngẫu nhiên nhiều lần trong trường hợp tính năng suất lao động (NSLĐ) chung của đơn vị Bưu điện Y có tổng số công nhân là 200 người, trong đó chọn ra 100 công nhân để điều tra thực tế về kết quả điều tra như sau:

Bảng 4.1

NSLĐ của một công nhân x_i	Số công nhân f_i	x'_i	$f_i x'_i$	$x'_i - \bar{x}$	$(x'_i - \bar{x})^2$	$f_i (x'_i - \bar{x})^2$
35 - 45	14	20	560	-18	324	4.536
45 - 55	20	50	1.000	-8	64	1.280
55 - 65	42	60	2.520	2	4	168
65 - 75	20	70	1.400	12	144	2.880
75 - 85	4	80	320	22	484	1.936
Cộng	100		5.800			10.800

Ta có:

$$\bar{x} = \frac{\sum x'_i f_i}{\sum f_i} = \frac{5.800}{100} = 58 \text{ (triệu đồng)}$$

$$\delta_m^2 = \frac{10.800}{100} = 108$$

$$\mu_x = \sqrt{\frac{108}{100-1}} = 1,044$$

b) Khi điều tra chọn mẫu để suy rộng tỷ lệ theo một tiêu thức nào đó thì sai số được tính như sau:

- Trường hợp chọn mẫu nhiều lần:

$$\mu_p = \sqrt{\frac{pq}{n}} = \sqrt{\frac{p(1-p)}{n}}$$

Trong đó: p - Tỷ lệ cấu thành của tổng thể chung:

$$p = \frac{M}{N}$$

M - Là số đơn vị mang dấu hiệu A nào đó mà ta quan tâm, theo định nghĩa xác suất thì p chính là xác suất để lấy đơn vị mang dấu hiệu A trong tổng thể chung.

q - Là tỷ trọng của bộ phận không mang dấu hiệu A của tổng thể chung.

Do tỷ lệ cấu thành chung không tính được nên phải dùng tỷ lệ cấu thành mẫu điều chỉnh để thay thế:

$$\mu_p = \sqrt{\frac{w(1-w)}{n-1}}$$

Trong đó: w - Tỷ lệ cấu thành của tổng thể mẫu $w = \frac{m}{n}$

m - Là số đơn vị mang dấu hiệu A nào đó mà ta quan tâm thuộc tổng thể mẫu.

n - Số đơn vị của tổng thể mẫu.

- Trong trường hợp chọn mẫu một lần

$$\mu_p = \sqrt{\frac{pq}{n} \left(1 - \frac{n}{N}\right)} = \sqrt{\frac{p(1-p)}{n} \left(1 - \frac{n}{N}\right)}$$

$$\mu_p = \sqrt{\frac{w(1-w)}{n}}$$

Ví dụ: Từ số liệu bảng 4.1, tính sai số trung bình chọn mẫu về tỷ lệ số công nhân có năng suất lao động từ 65 triệu đồng trở lên.

Ta có tỷ lệ số công nhân có năng suất lao động từ 65 triệu đồng trở lên là:

$$w = (20 + 4)/100 = 0,24$$

$$w(1 - w) = 0,24(1 - 0,24)$$

Do đó:

$$\mu_p = \sqrt{\frac{0,24(1 - 0,24)}{100 - 1}} = 0,00184$$

4.2.4.3. Ý nghĩa của μ_x và μ_p

Sai số μ_x và μ_p biểu hiện trị số của các sai số chọn mẫu gặp phải khi suy rộng tài liệu. Nhưng do chọn mẫu ngẫu nhiên nên sai số này không được xác định trước về dấu (+ hoặc -) mà nó phản ánh một phạm vi chênh lệch có thể nhiều hơn hay ít hơn so với

tham số của tổng thể chung, do đó chênh lệch giữa số trung bình mẫu (tỷ lệ mẫu) và số trung bình chung (tỷ lệ chung) là $\pm \mu_x$ và μ_p , nghĩa là:

$$\bar{X} = \bar{x} \pm \mu_x$$

$$p = w \pm \mu_p$$

$\bar{X}$ - Là số trung bình chung.

$\bar{x}$ - Số trung bình mẫu.

p - Tỷ lệ chung.

w - Tỷ lệ mẫu.

4.2.4.4. Phạm vi sai số chọn mẫu

Trong ví dụ trên ta tính được năng suất lao động trung bình (NSLĐ) mẫu là $\bar{x} = 58$ triệu, sai số chọn mẫu là $\mu_x = 1,044$. Như vậy có thể suy ra năng suất lao động trung bình của tất cả công nhân trong đơn vị Bưu điện Y là: $\bar{X} = 58 \pm 1,044$ (triệu đồng), nghĩa là trong phạm vi từ 56,956 triệu đến 59,044 triệu.

Theo chứng minh của toán học thì xác suất (hay trình độ tin cậy) của việc suy rộng kết quả nói trên mới chỉ bằng 0,6827, nghĩa là trong 10.000 trường hợp có 6.827 trường hợp chắc chắn sai số chọn mẫu không vượt quá $\pm \mu$.

Có thể nâng cao trình độ tin cậy của việc suy rộng tài liệu. Giả sử phạm vi này được mở rộng gấp hai lần (2μ) thì NSLĐ trung bình của đơn vị sẽ là:

$$\bar{X} = 58 \pm 2 \times 1,044 = 58 \pm 2,088$$

Nghĩa là trong phạm vi từ 55,912 triệu đến 60,088 triệu. Trong trường hợp này trình độ của việc suy rộng lên tới 0,9545. Tương tự nếu mở rộng phạm vi sai số chọn mẫu lên tới $\pm 3\mu$, thì trình độ tin cậy lên tới 0,9973.

Như vậy, càng mở rộng phạm vi sai số chọn mẫu thì trình độ tin cậy của việc suy rộng tài liệu càng tăng, đồng thời sai số chọn mẫu sẽ tăng theo. Hệ số dùng để mở rộng đó được gọi là hệ số tin cậy (t), nó ứng với hàm xác suất $\Phi(t)$ và đã được nhà toán học người Nga Lia-pu-nốp lập bảng tính sẵn.

Phạm vi sai số chọn mẫu có thể được xác định theo công thức:

$$\varepsilon = t\mu$$

Trong đó: ε - Phạm vi sai số chọn mẫu.

t - Hệ số tin cậy

μ - Sai số trung bình chọn mẫu.

Hoặc: $\varepsilon_x = t\mu_x$

$$\varepsilon_p = t\mu_p$$

4.2.4.5. Suy rộng kết quả điều tra chọn mẫu

Tức là tính toán các tham số của tổng thể chung trên cơ sở thu thập được trong điều tra chọn mẫu. Có hai phương pháp suy rộng:

a) Phương pháp tính đối trực tiếp

Được áp dụng khi dùng các số trung bình mẫu hay tỷ lệ mẫu của tổng thể mẫu để tính ra các tham số của tổng thể chung:

$$\bar{X} = \bar{x} \pm \varepsilon_x \Rightarrow$$

$$\bar{x} - \varepsilon_x \leq \bar{X} \leq \bar{x} + \varepsilon_x$$

$$p = w \pm \varepsilon_p$$

$$w - \varepsilon_p \leq p \leq w + \varepsilon_p$$

Ví dụ: Giả sử khi điều tra chọn mẫu, tỷ lệ sản phẩm dịch vụ đảm bảo chất lượng là 90%. Với xác suất 0,9545 tính ra phạm vi sai số chọn là $\varepsilon_p = 0,02$. Như vậy suy ra tỷ lệ sản phẩm dịch vụ đảm bảo chất lượng sẽ nằm trong phạm vi:

$$w - \varepsilon_p \leq p \leq w + \varepsilon_p$$

$$0,90 - 0,02 \leq p \leq 0,90 + 0,02$$

$$0,88 \leq p \leq 0,92$$

b) Phương pháp hệ số điều chỉnh

Được dùng để xác định kết quả điều tra toàn bộ. Căn cứ vào kết quả điều tra toàn bộ và điều tra chọn mẫu, tính ra tỷ lệ chênh lệch rồi dùng tỷ lệ này để làm hệ số điều chỉnh kết quả điều tra toàn bộ.

Ví dụ: Trong tổng điều tra dân số, tiến hành điều tra chọn mẫu trên 5% số địa điểm điều tra với mục đích nghiên cứu thêm một số tiêu thức, đồng thời kiểm tra kết quả tổng điều tra dân số. Giả sử theo kết quả tổng điều tra dân số thì số dân của tỉnh Y là 1.765.200 người, trong đó chỉ tính riêng 5% địa bàn điều tra (là những địa bàn chọn làm mẫu) là 95.260 người. Trong khi đó kết quả điều tra chọn mẫu trên 5% địa bàn thấy điều tra toàn bộ bỏ sót 48 người trong các địa bàn đó. Do đó tỷ lệ bỏ sót là: $48/95.260 = 0,0005$ hay 0,05%.

Dùng tỷ lệ này làm hệ số điều chỉnh kết quả điều tra toàn bộ. Như vậy trên toàn tỉnh số người bị bỏ sót có thể là:

$$0,0005 \times 1.765.200 = 883 \text{ người.}$$

Do đó số dân thực tế của tỉnh Y sẽ là:

$$1.765.200 + 883 = 1.766.083 \text{ người.}$$

4.2.4.6. Xác định số đơn vị tổng thể mẫu

Khi điều tra chọn mẫu cần xác định số đơn vị điều tra sao cho thỏa mãn hai yêu cầu sau:

- Bảo đảm sai số chọn mẫu nhỏ nhất;
- Chi phí thấp nhất.

Hai yêu cầu này đối lập nhau. Muốn đảm bảo tài liệu điều tra chính xác (sai số nhỏ nhất) phải điều tra một lượng đơn vị khá lớn, tức là cần một lượng kinh phí lớn. Ngược lại với một lượng kinh phí có hạn, chỉ có thể điều tra một số ít đơn vị thì phải chấp nhận một sai số nhất định.

Trong thực tế thường căn cứ vào yêu cầu của độ chính xác (phạm vi sai số chọn mẫu) để tính số đơn vị mẫu cần điều tra.

Trường hợp để ước lượng số trung bình:

- Nếu dự định chọn đơn vị tổng thể mẫu theo cách chọn nhiều lần (chọn hoàn lại):

$$\varepsilon_x = t\mu_x = t\sqrt{\frac{\delta^2}{n}} \quad \text{suy ra} \quad n = \frac{t^2\delta^2}{\varepsilon_x^2}$$

- Trong trường hợp chọn một lần (chọn không hoàn lại):

$$\varepsilon_x = t\sqrt{\frac{\sigma_x^2}{n} \left(1 - \frac{n}{N}\right)} \quad \text{suy ra} \quad n = \frac{Nt^2\delta^2}{N\varepsilon_x^2 + t^2\delta^2}$$

Trường hợp để ước lượng tỷ lệ:

- Nếu dự định chọn các đơn vị tổng thể theo cách chọn nhiều lần:

$$\varepsilon_r = t\mu_r = t\sqrt{\frac{pq}{n}} \quad \text{suy ra} \quad n = \frac{t^2 pq}{\varepsilon_r^2}$$

- Nếu dự định chọn mẫu theo cách chọn một lần:

$$\varepsilon_r = t\mu_r = t\sqrt{\frac{pq}{n} \left(1 - \frac{n}{N}\right)} \quad \text{suy ra} \quad n = \frac{Nt^2 pq}{N\varepsilon_r^2 + t^2 pq}$$

Qua các công thức trên cho thấy các nhân tố ảnh hưởng đến số đơn vị mẫu điều tra bao gồm:

- Phạm vi sai số chọn mẫu. Nếu yêu cầu phạm vi sai số càng nhỏ thì số đơn vị mẫu cần điều tra càng nhiều và ngược lại. Số đơn vị mẫu tỷ lệ nghịch với bình phương phạm vi sai số chọn mẫu.

- Hệ số tin cậy (t). Yêu cầu trình độ tin cậy của tài liệu suy rộng cao hay thấp thì hệ số tin cậy cũng phải lớn hay nhỏ một cách tương ứng và do đó số đơn vị mẫu cũng phải tăng giảm cho phù hợp. Số đơn vị mẫu tỷ lệ thuận với bình phương hệ số tin cậy.

- Tính chất đồng đều của hiện tượng nghiên cứu. Số đơn vị mẫu tỷ lệ thuận với phương sai của tổng thể chung δ^2 và pq. Nếu tiêu thức biến thiên ít, phương sai tính ra nhỏ, số đơn vị cần điều tra sẽ giảm đi. Ngược lại, nếu tiêu thức biến thiên lớn, phương sai sẽ lớn làm cho số đơn vị cần điều tra sẽ tăng lên.

Trong thực tế, khi tính số đơn vị mẫu cần điều tra thường không có tài liệu để tính phương sai chung, vì vậy phải giải quyết bằng cách lấy phương sai lớn nhất của các lần điều tra trước hoặc

lấy phương sai của hiện tượng tương tự, hoặc tổ chức điều tra thí điểm để lấy phương sai.

Ví dụ: Tại một Bưu chính - Phát hành báo chí, dùng phương pháp điều tra chọn mẫu để ước lượng tỷ lệ sản phẩm dịch vụ vi phạm chất lượng (tốc độ, độ chính xác và an toàn). Phạm vi sai số chọn mẫu là 0,08 về giá trị tuyệt đối, độ tin cậy phải đạt mức 99,73%. Xác định số mẫu cần chọn nếu:

- Các cuộc điều tra tương tự cho tỷ lệ sản phẩm dịch vụ đảm bảo chất lượng là 5%, 6%, 10%.

- Không có tài liệu về các cuộc điều tra lần trước.

Theo điều kiện đặt ra, cần phải xác định số mẫu (số sản phẩm dịch vụ Công ty xử lý trong thời gian đó).

* Trường hợp sử dụng tài liệu của cuộc điều tra lần trước: Ta có 03 kết quả về tỷ lệ sản phẩm dịch vụ vi phạm chất lượng ở các cuộc điều tra lần trước là $p_1 = 0,05$; $p_2 = 0,06$; $p_3 = 0,10$. Ở đây sử dụng $p_3 = 0,10$ vì giá trị này gần 0,5 hơn; $\varepsilon = 0,08$; $2\Phi(t) = 0,9973$ suy ra $t = 3$.

Số mẫu cần thiết có thể xác định:

$$n = \frac{t^2 p(1-p)}{\varepsilon^2} = \frac{3^2 \times 0,10(1-0,10)}{0,08^2} = 126,56 \approx 127$$

* Trường hợp không có tài liệu về các cuộc điều tra lần trước: Phải giả thiết $p = 0,5$ để phương sai $pq = 0,25$ là lớn nhất. Khi đó số mẫu tính ra cũng lớn nhất, nhưng đảm bảo được tính đại diện của tổng thể mẫu cao hơn.

$$n = \frac{t^2 p(1-p)}{\varepsilon^2} = \frac{3^2 \times 0,50(1-0,50)}{0,08^2} = 351,56 \approx 352$$

Sau khi xác định được số đơn vị mẫu điều tra, người ta thường kiểm tra tính chất đại biểu của tổng thể đó. Đây là việc làm cần thiết vì nếu tổng thể mẫu không đại biểu cho tổng thể chung thì kết quả suy rộng sẽ không chính xác. Việc kiểm tra tính chất đại biểu của tổng thể mẫu có thể tiến hành bằng nhiều cách:

- So sánh một số chỉ tiêu trung bình hoặc chỉ tiêu tỷ lệ của tổng thể mẫu đã chọn với các chỉ tiêu tương ứng của tổng thể chung ($\bar{X}$ với $\bar{x}$, w với p). Tất nhiên là giữa những chỉ tiêu này thường có sự chênh lệch, nhưng nếu mức độ chênh lệch không vượt quá phạm vi cho phép ($\pm 5\%$) thì tổng thể mẫu được coi như đủ tính chất đại biểu. Nếu mức độ chênh lệch vượt quá phạm vi $\pm 5\%$, tổng thể mẫu chưa đủ tính chất đại biểu, do đó phải thay thế một số đơn vị hoặc tăng thêm số đơn vị điều tra.

- Kiểm tra và xử lý những số liệu nghi ngờ. Nếu trong quá trình chọn các đơn vị điều tra gặp phải một vài đơn vị có số liệu quá lớn hay quá bé thì cần kiểm tra và xử lý bằng cách:

+ Nếu số đơn vị có số liệu đột xuất không nhiều, thì nên loại bỏ thay thế bằng các đơn vị khác.

+ Nếu số đơn vị có số liệu đột xuất không phải là cá biệt, thì nên chọn phân loại để đảm bảo có một số mẫu nhất định đại diện cho tổng thể chung.

+ Có thể so sánh sự phân phối các đơn vị theo tiêu thức nghiên cứu của 02 tổng thể. Cách so sánh này phức tạp hơn nhưng đảm bảo tính đại biểu hơn.

4.2.5. Các phương pháp tổ chức chọn mẫu

4.2.5.1. Phương pháp chọn mẫu ngẫu nhiên đơn thuần

Đây là phương pháp tổ chức chọn các đơn vị mẫu trong tổng thể chung một cách hết sức ngẫu nhiên không qua một sự sắp xếp nào. Mỗi đơn vị của tổng thể chung có thể được chọn một lần hoặc nhiều lần.

Phương pháp này đơn giản nhưng nó chỉ cho kết quả tốt khi tổng thể nghiên cứu tương đối đồng đều. Nếu tổng thể nghiên cứu có kết cấu phức tạp thì không nên chọn phương pháp này vì chọn như vậy số đơn vị tổng thể mẫu sẽ không được phân phối đều trong tổng thể chung, tính chất đại biểu của tổng thể mẫu sẽ không cao. Nếu tổng thể chung có qui mô lớn thì việc đặt số liệu rút thăm cho tất cả các đơn vị cũng gặp khó khăn. Chính vì vậy trong thực tế phương pháp chọn ngẫu nhiên đơn thuần ít được sử dụng.

4.2.5.2. Phương pháp chọn máy móc

Đây là phương pháp tổ chức chọn mẫu trong đó mỗi đơn vị được chọn căn cứ vào những khoảng cách nhất định. Trước hết xếp các đơn vị tổng thể chung theo một thứ tự nào đó như sắp xếp theo vần A, B, C... của tên gọi, theo thứ tự, theo qui mô từ nhỏ tới lớn, v.v... Sau đó lần lượt chọn các đơn vị theo thứ tự một cách máy móc, tức là cứ sau một khoảng cách nhất định lại chọn ra một đơn vị. Khoảng cách này được xác định bằng cách lấy số đơn vị tổng thể chung chia cho số đơn vị tổng thể mẫu:

$$d = \frac{N}{n}$$

Ví dụ: Cần chọn 200 công nhân trong số 2.000 công nhân của Bưu điện tỉnh X để điều tra năng suất lao động. Có thể sắp

xếp 2.000 công nhân đó thành danh sách theo vần của tên gọi (A, B, C...), lúc đó khoảng cách để chọn ra từng người là $2000/200 = 10$. Tức là cứ cách 10 người trong danh sách lại chọn ra một người. Người đầu tiên được chọn ra trong số 10 người đầu của danh bằng cách ngẫu nhiên đơn thuần. Giả sử trong 10 người này rút thăm được người thứ 5, thì những người được chọn tiếp theo là 15, 25...

Trong thực tế có những hiện tượng mà các đơn vị của nó đã được sắp xếp sẵn và trong suốt thời gian điều tra vẫn không thay đổi vị trí. Với những hiện tượng như vậy có thể lợi dụng sự bố trí sẵn có để chọn theo khoảng cách đã tính.

Phương pháp chọn máy móc có thủ tục tiến hành đơn giản hơn chọn ngẫu nhiên đơn thuần. Do lựa chọn theo từng khoảng cách nhất định nên số đơn vị mẫu được phân đều trong tổng thể chung, tính chất đại biểu của tổng thể mẫu được nâng cao.

4.2.5.3. Phương pháp chọn phân loại

Đó là tiến hành chọn các đơn vị mẫu khi tổng thể chung đã được phân chia thành các tổ theo các tiêu thức liên quan trực tiếp đến mục đích nghiên cứu. Việc chọn các đơn vị từ các tổ được tiến hành theo phương pháp chọn mẫu ngẫu nhiên.

Khi phân tổ, các tổ thường có các quy mô khác nhau. Nếu chọn số đơn vị tổng thể mẫu trong mỗi tổ tương ứng với tỷ trọng của nó chiếm trong tổng thể thì gọi là chọn phân loại theo tỷ lệ. Nếu chọn số đơn vị mẫu trong các tổ đều bằng nhau thì gọi là chọn không theo tỷ lệ. Trong chọn tỷ lệ, số đơn vị chọn trong mỗi

tổ được tính theo công thức:
$$n_i = n \frac{N_i}{N}$$

Đối với phân loại tỷ lệ sai số trung bình chọn mẫu không phụ thuộc vào phương sai chung mà phụ thuộc vào trung bình các phương sai tổ. Vì vậy có công thức tính sau:

Khi nhiệm vụ là để suy rộng số trung bình:

- Trường hợp chọn nhiều lần:

$$\mu_x = \sqrt{\frac{\delta^2}{n}}$$

- Trường hợp chọn một lần:

$$\mu_x = \sqrt{\frac{\delta^2}{n} \left(1 - \frac{n}{N}\right)}$$

Với:

$$\delta^2_x = \frac{\sum \delta_i N_i}{\sum N_i}$$

Khi nhiệm vụ là để ước lượng tỷ lệ:

- Trường hợp chọn nhiều lần:

$$\mu_p = \sqrt{\frac{p(1-p)}{n}}$$

- Trường hợp chọn một lần:

$$\mu_p = \sqrt{\frac{p(1-p)}{n} \left(1 - \frac{n}{N}\right)}$$

Với:

$$p(1-p) = \frac{\sum p(1-p) N_i}{\sum N_i}$$

Sai số trung bình chọn mẫu trong chọn phân loại nhỏ hơn trong chọn mẫu ngẫu nhiên đơn thuần và máy móc.

Đối với chọn phân loại không tỷ lệ, sai số trung bình chọn mẫu được tính như sau:

$$\mu = \frac{1}{N} \sqrt{\mu_i^2 N_i}$$

Trong đó: μ_i - Sai số trung bình trong từng tổ;

N_i - Số đơn vị trong từng tổ của tổng thể chung.

4.2.5.4. Phương pháp chọn cả khối

Theo phương pháp này số mẫu được rút ra không phải là lẻ tẻ đơn vị mà từng nhóm (khối đơn vị). Mỗi nhóm đơn vị được rút ra được điều tra hết không bỏ sót một đơn vị nào. Có thể nói rằng đây là điều tra toàn bộ trong các khối được chọn ra.

Muốn chọn theo phương pháp này, trước hết tổng thể chung phải được chia thành các khối có qui mô bằng nhau hoặc khác nhau (để tiện cho việc tính toán thường chia các khối có qui mô bằng nhau). Việc chọn khối có thể bằng phương pháp chọn ngẫu nhiên đơn thuần hoặc chọn máy móc.

Trong chọn cả khối sai số trung bình chọn mẫu không phụ thuộc vào phương sai chung mà phụ thuộc vào sự chênh lệch giữa các số trung bình khối hay còn gọi là phương sai các số trung bình khối. Vì vậy sai số trung bình chọn mẫu được tính theo công thức sau:

- Khi suy rộng số trung bình:

$$\mu_x = \sqrt{\frac{\delta}{r} \left(\frac{R-r}{R-1} \right)}$$

- Khi suy rộng tỷ lệ:

$$\mu_p = \sqrt{\frac{w(1-w)}{r} \left(\frac{R-r}{R-1} \right)}$$

Trong đó: R - Số khối của tổng thể chung;

r - Số khối được chọn ra để điều tra.

$$\delta^2 = \frac{\sum (\bar{x}_i - \bar{x})^2 n_i}{\sum n_i}$$

Trong đó: $\bar{x}_i$ - Số trung bình của khối thứ i được chọn;

$\bar{x}$ - Số trung bình của tất cả các khối được chọn.

$$w(1-w) = \frac{\sum (w_i - w_r)^2 n_i}{\sum n_i}$$

$$w_r = \frac{\sum w_i n_i}{\sum n_i}$$

Nếu số đơn vị các khối bằng nhau:

$$w_r = \frac{\sum w_i}{r}$$

Trong đó: w_i - Tỷ lệ từng khối được chọn;

w_r - Tỷ lệ trung bình các khối được chọn.

4.2.5. Phương pháp chọn kết hợp

Chọn kết hợp là sử dụng kết hợp một số phương pháp chọn với nhau. Nếu chọn kết hợp cả khối với chọn ngẫu nhiên đơn

thuận thì trước hết tổng thể chung được chia thành các khối, chọn một số khối cần thiết, sau đó chọn ngẫu nhiên các đơn vị trong các khối đã chọn để điều tra. Với cách chọn này, sai số trung bình chọn mẫu được tính theo công thức sau:

Trường hợp chọn nhiều lần:

$$\mu_x^2 = \frac{\delta^2}{n} + \frac{\delta^2}{r}$$

Trường hợp chọn một lần:

$$\mu_x^2 = \frac{\delta^2}{n} \left(1 - \frac{n}{N} \right) + \frac{\delta^2}{r} \left(1 - \frac{r}{R} \right)$$

4.2.6. Điều tra chọn mẫu nhỏ và điều tra chọn mẫu thời điểm

4.2.6.1. Điều tra chọn mẫu nhỏ

Trên thực tế có nhiều trường hợp không thể điều tra một số đơn vị tương đối lớn vì nó liên quan đến việc phá hủy đơn vị điều tra như điều tra chất lượng đồ hộp, thử độ bền của bóng đèn v.v.... Vì vậy nảy sinh yêu cầu chọn mẫu nhỏ, nghĩa là tìm hiểu đặc điểm của tổng thể chung từ một tổng thể mẫu nhỏ, với số mẫu thường không quá hai mươi đơn vị. Trong điều tra chọn mẫu nhỏ, nếu dùng phương pháp thích hợp vẫn có thể đảm bảo độ chính xác để suy rộng tài liệu.

Quan hệ giữa phương sai mẫu và phương sai chung có thể biểu hiện:

$$\delta^2 = \frac{n}{n-1} \delta_m^2$$

Trong điều tra chọn mẫu nhỏ (với $n < 20$) tỷ số $\frac{n}{n-1}$ là đáng kể. Vì vậy công thức tính sai số trung bình chọn mẫu trong điều tra chọn mẫu nhỏ là:

$$\mu_0 = \sqrt{\frac{\delta^2}{n}} = \sqrt{\frac{\delta_m^2}{n-1}}$$

$$\mu_0 = \sqrt{\frac{w(1-w)}{n-1}}$$

Trong đó: μ_0 - Sai số trung bình chọn mẫu nhỏ

$$\text{Và } \varepsilon_0 = t\mu_0$$

Như vậy hệ số tin cậy t càng cao thì phạm vi sai số chọn mẫu càng lớn và xác suất suy rộng càng tăng. Khác với các cuộc điều tra chọn mẫu bình thường, ở đây xác suất ứng với các hệ số tin cậy t được tính trong bảng phân phối Student.

Ví dụ: Để kiểm tra chất lượng sản phẩm, người ta chọn ngẫu nhiên 16 hộp để kiểm tra. Kết quả kiểm tra cho thấy trong 16 hộp có 02 hộp kém phẩm chất.

Vậy tỷ lệ sản phẩm kém chất lượng:

$$w = \frac{2}{16} = 0,125$$

Sai số trung bình chọn mẫu nhỏ:

$$\mu_0 = \sqrt{\frac{w(1-w)}{n-1}} = \sqrt{\frac{0,125(1-0,125)}{16-1}} = 0,00729$$

Nếu suy rộng tỷ lệ sản phẩm kém phẩm chất cho toàn bộ sản phẩm với xác suất 0,936, ta có $n = 16$ thì $t = 2$ do đó:

$$\varepsilon_0 = t\mu_0 = 2 \times 0,00729 = 0,01458$$

Vậy tỷ lệ phế phẩm: $0,125 - 0,01458 < p < 0,125 + 0,01458$

Hay: $0,1104 < p < 0,1396$

4.2.6.2. Điều tra chọn mẫu thời điểm

Điều tra chọn mẫu thời điểm thường được dùng trong công nghiệp và giao thông vận tải để nghiên cứu thời gian làm việc của công nhân hoặc của thiết bị. Đây là một phương pháp chọn mẫu đặc biệt vì xét theo thời gian thì nói là chọn mẫu, nhưng xét theo phạm vi tổng thể nghiên cứu thì lại là điều tra toàn bộ. Nội dung của phương pháp là trong những thời điểm nhất định người ta đăng ký sự tồn tại của các phần tử thuộc quá trình nghiên cứu, không kể thời gian tồn tại đó dài hay ngắn.

Ví dụ: Nghiên cứu thời gian làm việc của công nhân, người ta chia thành hai phần: Làm việc và ngừng việc. Khi đi kiểm tra tại thời điểm (làm việc hay ngừng việc) giả sử có 40 công nhân làm việc cứ cách 30 phút lại đi kiểm tra một lần. Trong 8 giờ làm việc đã đăng ký được $8 \times 2 \times 40 = 640$ trường hợp. Trong đó có 576 trường hợp làm việc, 64 trường hợp ngừng việc. Như vậy tỷ lệ công nhân làm việc là 0,9. Với hệ số tin cậy $t = 2$, phạm vi sai số chọn mẫu là:

$$\varepsilon_p = \sqrt{\frac{0,9 \times 0,1}{640}} = 0,0236$$

Có nghĩa là với xác suất 0,9545 có thể xác định tỷ lệ sử dụng thời gian làm việc của công nhân trong cả ca làm việc là $0,9 - 0,0236 < p < 0,9 + 0,0236$ hay $0,8764 < p < 0,9236$.

4.3. ĐIỀU TRA CHỌN MẪU PHI NGẪU NHIÊN

Điều tra chọn mẫu phi ngẫu nhiên là chọn có dụng ý trước, nghĩa là dựa trên sự hiểu biết về hiện tượng nghiên cứu, tiến hành bàn bạc, phân tích để lựa chọn những đơn vị điển hình có khả năng đại diện cho tổng thể nghiên cứu để điều tra. Kết quả điều tra thường được dùng để suy rộng cho toàn bộ tổng thể hoặc để đánh giá một cách tổng quát. Điều tra chọn mẫu phi ngẫu nhiên được sử dụng rộng rãi để nghiên cứu các hiện tượng kinh tế phức tạp, phân tán, không ổn định đòi hỏi phải quan sát, phân tích tỷ mỉ trước khi thu thập tài liệu.

Điều tra chọn mẫu phi ngẫu nhiên không hoàn toàn dựa trên cơ sở toán học như chọn mẫu ngẫu nhiên mà chủ yếu đòi hỏi phải kết hợp chặt chẽ giữa phân tích lý luận và thực tế. Sự nhận xét chủ quan của người tổ chức có ảnh hưởng đến chất lượng điều tra. Muốn đảm bảo chất lượng tài liệu điều tra phải giải quyết các vấn đề sau:

4.3.1. Đảm bảo phân tổ chính xác đối tượng điều tra

Vì mỗi đơn vị được chọn ra dù có đầy đủ tính chất đại biểu đến mấy cũng chỉ có khả năng đại diện cho mỗi bộ phận, một loại hình nào đó trong tổng thể phức tạp. Nếu tập hợp được các điển hình của nhiều bộ phận thì chúng có khả năng đại diện cho cả tổng thể hiện tượng phức tạp. Mặt khác, việc phân tổ có tác

dụng thu hẹp độ biến thiên tiêu thức trong mỗi bộ phận làm cho việc suy rộng tài liệu càng tỷ mỉ và chính xác.

Tổng thể càng phức tạp, việc phân tích càng phải thận trọng, nhiều khi phân tổ còn phải trải qua nhiều bước để có những tổ chi tiết hơn.

4.3.2. Chọn đơn vị điều tra

Trong điều tra chọn mẫu phi ngẫu nhiên, người ta chọn các đơn vị điển hình có khả năng đại diện cho từng bộ phận khác nhau của tổng thể nghiên cứu. Có nhiều cách chọn:

- Chọn những đơn vị có mức độ tiêu thức gần với số trung bình của từng bộ phận nhất, đồng thời cũng là mức độ phổ biến nhất trong bộ phận đó. Khi chọn phải thông qua quan sát bàn bạc phân tích tập thể thì mới chọn được những đơn vị điều tra có tính đại biểu cao. Sau khi chọn được các đơn vị điều tra phải kiểm tra tính chất đại biểu của các đơn vị đó, nếu chấp nhận được thì mới tiến hành điều tra thực tế.

- Chọn những đơn vị có kinh nghiệm về một mặt nào đó.

- Chọn một số địa phương đại diện cho từng vùng kinh tế. Trong các vùng này lại chọn ngẫu nhiên một số đơn vị điều tra.

4.3.3. Xác định số đơn vị điều tra

Điều tra chọn mẫu phi ngẫu nhiên cũng phải dựa trên cơ sở định luật số lớn, nghĩa là cần chọn ra một số đơn vị điều tra nhiều tới mức đủ khả năng đại biểu cho cả tổng thể. Vì chọn phi ngẫu nhiên nên không thể dùng công thức toán học để tính. Muốn xác định số mẫu cần:

- Căn cứ vào tính chất phức tạp của tổng thể điều tra. Tổng thể càng phức tạp càng cần điều tra nhiều đơn vị.

- Căn cứ vào kinh nghiệm của các địa phương khác, nước khác, các lần điều tra trước để quyết định số đơn vị cần điều tra thực tế.

- Căn cứ vào mức độ đòi hỏi của việc nghiên cứu, lực lượng cán bộ và khả năng vật chất để quyết định tăng thêm hay giảm bớt số đơn vị cần điều tra.

4.3.4. Sai số chọn mẫu

Trong điều tra chọn mẫu phi ngẫu nhiên, sai số chọn mẫu không thể tính được bằng công thức toán học, mà phải thông qua nhận xét, so sánh để ước lượng ra. Nếu thấy sai số không lớn lắm có thể chấp nhận được thì dùng kết quả của điều tra chọn mẫu suy rộng ra các tham số của tổng thể chung. Nếu thấy nghi ngờ thì có thể chọn lại và điều tra lại. Ngoài ra có thể dùng phương pháp kiểm định thống kê để xác định chất lượng kết quả điều tra chọn mẫu.

Khi suy rộng thì suy rộng trực tiếp, chứ không suy rộng có phạm vi như trong điều tra chọn mẫu ngẫu nhiên. Các đơn vị điều tra được lựa chọn từ các bộ phận khác nhau nên khi suy rộng phải theo thứ tự từng bước và chú ý đến tỷ trọng của mỗi bộ phận trong tổng thể.

4.3.5. Lực lượng tham gia điều tra

Điều tra chọn mẫu phi ngẫu nhiên đòi hỏi phải giải quyết nhiều vấn đề phức tạp, mà cơ sở chủ yếu là dựa trên sự phân tích sâu sắc đối tượng nghiên cứu. Muốn làm tốt công tác điều tra, phải thành thạo về nghiệp vụ, am hiểu hiện tượng nghiên cứu, làm tốt công tác tổ chức vận động để hiểu rõ mục đích nghiên cứu, giải quyết tốt những vấn đề ảnh hưởng tới chất lượng điều tra.

Chương 5

TƯƠNG QUAN VÀ HỒI QUY

5.1. KHÁI NIỆM

5.1.1. Tiêu thức nguyên nhân và tiêu thức kết quả

Giữa các hiện tượng kinh tế xã hội tồn tại những mối liên quan ràng buộc lẫn nhau. Ngay trong cùng một hiện tượng nghiên cứu bao gồm nhiều tiêu thức khác nhau, thì những tiêu thức này cũng có những mối liên hệ qua lại nhất định. Nếu xét về mức độ phụ thuộc giữa các tiêu thức thì người ta thường chia mối liên hệ giữa các hiện tượng thành các loại sau:

Liên hệ hàm số: Là loại liên hệ hết sức chặt chẽ giữa hai đại lượng nghiên cứu và biểu hiện sự liên hệ này theo những tỷ lệ tương ứng chặt chẽ, tức là khi hiện tượng thay đổi thì nó hoàn toàn quyết định sự thay đổi của hiện tượng liên quan theo một tỷ lệ tương ứng. Trường hợp này ít gặp trong thực tế, thường gặp trong toán học và vật lý.

Liên hệ tương quan: Là mối liên hệ không hoàn toàn chặt chẽ như liên hệ hàm số, tức là khi trị số của tiêu thức nguyên nhân thay đổi nó dẫn đến tiêu thức kết quả thay đổi theo. Nhưng sự thay đổi của tiêu thức kết quả không hoàn toàn phụ thuộc vào sự thay đổi của tiêu thức nguyên nhân nói trên mà nó còn phụ thuộc vào các tiêu thức nguyên nhân khác, thống kê gọi mối liên hệ đó là mối liên hệ tương quan. Để phân tích và đánh giá mối liên hệ

tương quan giữa các tiêu thức người ta thường sử dụng phương pháp tương quan, trên cơ sở quan sát hiện tượng số lớn. Phương pháp tương quan thường được vận dụng trong trường hợp trị số của tiêu thức nào đó bị thay đổi do ảnh hưởng của nhiều tiêu thức khác, trong đó mức độ ảnh hưởng của các tiêu thức có thể khác nhau. Tùy theo mục đích nghiên cứu có thể chọn riêng ra 1, 2 hoặc 3, 4 tiêu thức có ý nghĩa quyết định để xác định mối liên hệ tương quan giữa chúng với nhau. Trong thống kê thường phân biệt:

- *Tiêu thức nguyên nhân*: Là loại tiêu thức mà sự thay đổi trị số của nó là nguyên nhân làm ảnh hưởng tới sự thay đổi trị số của tiêu thức khác (thường ký hiệu x - là tiêu thức nguyên nhân);

- *Tiêu thức kết quả*: Là loại tiêu thức chịu ảnh hưởng tác động của các tiêu thức nguyên nhân (trong thống kê ký hiệu y - tiêu thức kết quả).

Ứng với mỗi giá trị (trị số) x_i của tiêu thức x , ta có một trị số y_j của tiêu thức y . Số liệu ban đầu của hai tiêu thức này được trình bày dưới dạng bảng.

5.1.2. Nhiệm vụ của phương pháp hồi quy và tương quan

5.1.2.1. Khái niệm phương pháp hồi quy và tương quan

Hồi quy và tương quan là phương pháp toán học được vận dụng để phân tích mối liên hệ tương quan giữa các hiện tượng nghiên cứu gồm nhiều yếu tố, giữa các yếu tố có ảnh hưởng lẫn nhau, liên hệ tương quan với nhau.

5.1.2.2. Nhiệm vụ của phương pháp hồi quy và tương quan

* *Xác định phương trình hồi quy* nhằm biểu diễn mối liên hệ tương quan dưới dạng một hàm số bao gồm:

- Phân tích bản chất của hiện tượng;
 - Chọn dạng hàm số phù hợp với hiện tượng số lớn đã quan sát và bản chất của hiện tượng;
 - Tính toán các tham số của phương trình hồi quy.
- * *Đánh giá mức độ chặt chẽ của mối liên hệ tương quan bằng cách tính hệ số tương quan và tỷ số tương quan.*

5.2. LIÊN HỆ TƯƠNG QUAN TUYẾN TÍNH GIỮA HAI TIÊU THỨC SỐ LƯỢNG

Tương quan tuyến tính là tương quan biểu thị mối liên hệ giữa hai (hay nhiều) tiêu thức phù hợp với một phương trình đường thẳng nào đó. Trong trường hợp này các giá trị thực nghiệm của tiêu thức nguyên nhân (x) và tiêu thức kết quả (y) thường được trình bày dưới dạng bảng:

Bảng 5.1

x	x ₁	x ₂	...	x _i	...	x _n
y	y ₁	y ₂	...	y _i	...	y _n

Giả sử giữa hai tiêu thức này có mối liên hệ tương quan tuyến tính theo một phương trình đường thẳng:

$$y_x = a + bx$$

Trong đó: x - Giá trị thực nghiệm của tiêu thức nguyên nhân;

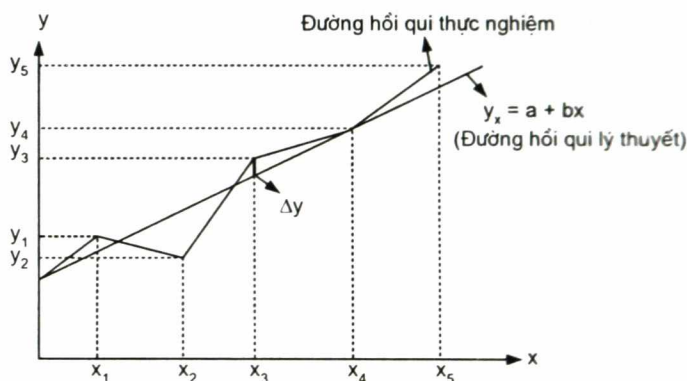
y_x - Giá trị lý thuyết của tiêu thức kết quả tính theo phương trình hồi quy;

a, b - Là các tham số qui định vị trí của phương trình hồi qui.

Quá trình xác định và đánh giá mối tương quan tuyến tính giữa hai tiêu thức số lượng được tiến hành theo trình tự sau:

Phương trình hồi qui:

Để phán đoán và lựa chọn dạng phương trình hồi qui cho phù hợp phải căn cứ vào các giá trị thực nghiệm của tiêu thức nguyên nhân x và tiêu thức kết quả y . Thông thường vẽ đồ thị các điểm thực nghiệm, mỗi điểm ứng với một cặp giá trị (x_i, y_i) , rồi từ đó bằng trực quan phán đoán dạng phương trình hồi qui thích hợp căn cứ và tình hình phân bố các điểm trên đồ thị.



Nếu các điểm thực nghiệm phân bố theo một dãy dài thì có thể chọn phương trình hồi quy (lý thuyết) dạng đường thẳng:

$$y_x = a + bx$$

Xác định phương trình này tức là phải tính được các tham số a , b . Thông thường các tham số a , b được xác định bằng phương pháp tổng bình phương nhỏ nhất. Đường hồi quy lý thuyết phải đảm bảo sao cho gần các điểm thực nghiệm nhất, nghĩa là chênh lệch giữa y_x và giá trị thực nghiệm y_i phải nhỏ nhất. Nếu gọi Δy_i là chênh lệch giữa giá trị thực nghiệm và giá trị lý thuyết thì:

Ứng với x_1 ta có $\Delta y_1 = y_1 - y_1(x)$

Ứng với x_2 ta có $\Delta y_2 = y_2 - y_2(x)$

...

Ứng với x_i ta có $\Delta y_i = y_i - y_i(x)$

...

Ứng với x_n ta có $\Delta y_n = y_n - y_n(x)$

Dấu của Δy_i có thể âm hoặc dương, do đó để khắc phục về dấu khi phải cộng các Δy_i , không lập tổng $\Sigma \Delta y_i$ mà lập tổng bình phương của các chênh lệch đó và yêu cầu tổng đó phải bé nhất, nghĩa là:

$$S = \Sigma \Delta y_i^2 = \Sigma (y - y_x)^2 \rightarrow \min$$

hay
$$S = \Sigma [y - (a + bx)]^2 \rightarrow \min$$

Muốn vậy đạo hàm riêng của S theo các tham số a, b phải triệt tiêu, nghĩa là:

$$\begin{cases} \frac{\partial S}{\partial a} = 0 \\ \frac{\partial S}{\partial b} = 0 \end{cases}$$

Hay:
$$\begin{cases} na + b \Sigma x = \Sigma y \\ a \Sigma x + b \Sigma x^2 = \Sigma xy \end{cases}$$

Hệ phương trình trên gọi là hệ phương trình chuẩn tắc xác định các tham số của phương trình hồi qui.

Giải hệ phương trình trên tính được a, b theo công thức sau:

$$b = \frac{\sum_i xy - \sum_i x \sum_i y}{\sum_i x^2 - \sum_i x \sum_i x} = \frac{\overline{xy} - \bar{x} \bar{y}}{\overline{x^2} - \bar{x}^2} = \frac{\overline{xy} - \bar{x} \bar{y}}{\delta_x^2}$$

$$a = \bar{y} - b\bar{x}$$

Thay các giá trị a , b vào phương trình hồi qui, ta sẽ có phương trình hồi qui biểu diễn mối liên hệ tương quan giữa hai tiêu thức số lượng x , y . Trong đó:

- a là hệ số tự do, phản ánh ảnh hưởng của các tiêu thức khác đến tiêu thức kết quả y ngoài tiêu thức x .

- b là hệ số hồi qui, phản ánh mức độ ảnh hưởng của tiêu thức nguyên nhân x đến tiêu thức kết quả y .

Hệ số tương quan:

Hệ số tương quan là một số tương đối (biểu hiện bằng số lần) dùng để đánh giá trình độ chặt chẽ của mối liên hệ tương quan tuyến tính. Hệ số này được tính từ các số liệu thực nghiệm của tiêu thức x và y . Căn cứ vào các giá trị tính được của nó có thể đánh giá được mối liên hệ tương quan tuyến tính có chặt chẽ hay không và chặt chẽ tới mức độ nào?

Hệ số tương quan có thể được tính theo công thức sau:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

Biến đổi ta được:

$$r = \frac{\overline{xy} - \bar{x}\bar{y}}{\delta_x \delta_y} = b \cdot \frac{\delta_x}{\delta_y}$$

Trong đó: x_i , y_i - Là các trị số thực nghiệm của tiêu thức nguyên nhân và tiêu thức kết quả;

$\bar{x}$, $\bar{y}$ - Là số trung bình cộng của các trị số x_i , y_i

δ_x , δ_y - Độ lệch tiêu chuẩn theo các tiêu thức x và y .

Giá trị của hệ số tương quan r luôn nằm trong khoảng $-1 \leq r \leq 1$.

+ Khi $r = +1$ hoặc $r = -1$ thì liên hệ giữa tiêu thức x , y là liên hệ hàm số.

+ Khi $r = 0$ thì giữa hai tiêu thức x và y không có mối liên hệ tương quan.

+ Khi r càng gần giá trị ± 1 thì mối liên hệ giữa tiêu thức x và y càng chặt chẽ và ngược lại.

+ Khi $1 > r > 0$ giữa x và y có mối tương quan tỉ lệ thuận.

+ Khi $-1 < r < 0$ giữa x và y có mối liên hệ tương quan tỉ lệ nghịch.

Để phản ánh đúng đắn mối liên hệ tương quan đòi hỏi phải nghiên cứu hiện tượng số lớn - tức là nghiên cứu nhiều đơn vị. Khi đó tài liệu thường được phân tổ kết hợp theo tiêu thức nguyên nhân (x) và tiêu thức kết quả (y). Việc phân tổ kết hợp sẽ hình thành bảng tương quan có dạng sau đây:

Bảng 5.2

x y		n_y
	n_{xy}	
n_x		N

Trong đó: n_x - Tần số các tổ được phân tổ theo tiêu thức x

n_y - Tần số các tổ được phân tổ theo tiêu thức y

n_{xy} - Tần số các tổ được phân tổ kết hợp theo tiêu thức x và tiêu thức y

N - Số đơn vị nghiên cứu

$$N = \sum n_x = \sum n_y = \sum \sum n_{xy}$$

Từ bảng tương quan, khi tính a, b, r phải nhân với các tần số tương ứng

$$\begin{cases} Na + b \sum x n_x = \sum n_y y \\ a \sum x n_x + b \sum x^2 n_x = \sum x y n_{xy} \end{cases}$$

Với công thức tính hệ số tương quan:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y}) n_{xy}}{\sqrt{\sum (x_i - \bar{x})^2 n_x \sum (y_i - \bar{y})^2 n_y}}$$

5.3. LIÊN HỆ TƯƠNG QUAN PHI TUYẾN GIỮA HAI TIÊU THỨC SỐ LƯỢNG

Tương quan phi tuyến là tương quan biểu thị mối quan hệ giữa các tiêu thức phù hợp với một đường cong nhất định (Parabol, Hypebol, hàm số mũ, loga...).

Khi giữa hai tiêu thức x và y tồn tại mối liên hệ tương quan phi tuyến thì phương trình đặc trưng của mối liên hệ tương quan này không phải là một đường thẳng mà là một dạng đường cong

nào đó (Parabol, Hypebol, hàm số mũ, loga...). Cách thức chung để xác định dạng phương trình hồi qui dạng phi tuyến cũng tương tự như khi xác định mối liên hệ tương quan tuyến tính.

5.3.1. Các dạng phương trình hồi qui

5.3.1.1. Phương trình Parabol bậc 2

Phương trình Parabol bậc 2 thường được sử dụng khi các giá trị của tiêu thức nguyên nhân tăng lên thì các trị số của tiêu thức kết quả tăng (hoặc giảm), việc tăng (hoặc giảm) đạt đến trị số cực đại (hoặc cực tiểu) rồi sau đó giảm (hoặc tăng). Phương trình hồi qui có dạng: $y_x = a + bx + cx^2$

Trong đó: a, b, c là các tham số, được xác định bằng phương pháp bình phương nhỏ nhất, bằng cách giải hệ phương trình:

$$\begin{cases} na + b\sum x + c\sum x^2 = \sum y \\ a\sum x + b\sum x^2 + c\sum x^3 = \sum xy \\ a\sum x^2 + b\sum x^3 + c\sum x^4 = \sum yx^2 \end{cases}$$

5.3.1.2. Phương trình Hypebol

Phương trình Hypebol được áp dụng trong trường hợp khi các trị số của tiêu thức nguyên nhân tăng lên thì các trị số của tiêu thức kết quả có thể giảm và đến một giới hạn nào đó ($\bar{y}_x = a$) thì hầu như không giảm. Phương trình hồi qui có

dạng: $y_x = a + \frac{b}{x}$

Các tham số a, b được xác định bằng phương pháp bình phương nhỏ nhất bằng cách giải hệ phương trình chuẩn tắc:

$$\begin{cases} na + b\sum \frac{1}{x} = \sum y \\ a\sum \frac{1}{x} + b\sum \frac{1}{x^2} = \sum \frac{y}{x} \end{cases}$$

5.3.1.3. Phương trình hàm mũ

Phương trình hàm mũ được áp dụng trong trường hợp cùng với sự tăng lên của các trị số tiêu thức nguyên nhân thì các trị số của các tiêu thức kết quả thay đổi theo cấp số nhân, nghĩa là có tốc độ phát triển xấp xỉ nhau. Phương trình hồi quy có dạng:

$$y_x = a b^x$$

Các tham số a , b được xác định bằng cách giải hệ phương trình chuẩn tắc:

$$\begin{cases} n \lg a + \lg b \sum x = \sum \lg y \\ \lg a \sum x + \lg b \sum x^2 = \sum (\lg y)x \end{cases}$$

Ngoài 03 dạng phương trình phi tuyến ở trên, còn có nhiều dạng khác như Parabol bậc 3, lũy thừa, compec...

5.3.2. Các chỉ tiêu đánh giá mối liên hệ tương quan phi tuyến tính

Sau khi tìm được phương trình hồi qui, ta đi đánh giá trình độ chặt chẽ của mối liên hệ tương quan phi tuyến. Thông thường để đánh giá trình độ chặt chẽ của mối liên hệ tương quan phi tuyến bằng tỷ số tương quan.

Khi có mối liên hệ tương quan giữa tiêu thức nguyên nhân x và tiêu thức kết quả y thì có thể tính các loại phương sai sau đây:

- Phương sai chung, phản ánh độ biến thiên của y do tất cả các tiêu thức nguyên nhân, trong đó có cả tiêu thức x :

$$\delta_y^2 = \frac{\sum (y_i - \bar{y})^2}{n}$$

- Phương sai phản ánh độ biến thiên của y do ảnh hưởng riêng của tiêu thức x :

$$\delta_{y_x}^2 = \frac{\sum (y_x - \bar{y})^2}{n}$$

- Phương sai phản ánh độ biến thiên của y do ảnh hưởng của các tiêu thức nguyên nhân khác trừ tiêu thức x :

$$\delta_{y_{(x)}}^2 = \frac{\sum (y_i - y_x)^2}{n}$$

Mối quan hệ giữa 03 phương sai:

$$\delta_y^2 = \delta_{y_x}^2 + \delta_{y_{(x)}}^2$$

Như vậy ta thấy nếu tiêu thức nguyên nhân x ảnh hưởng đến tiêu thức kết quả y càng lớn thì tỷ trọng của phương sai $\delta_{y_x}^2$, phản ánh độ biến thiên của tiêu thức y do ảnh hưởng của tiêu thức x chiếm trong phương sai chung δ_y^2 càng nhiều và ngược lại. Do đó tỷ số giữa phương sai $\delta_{y_x}^2$ và phương sai chung δ_y^2 có thể dùng để đánh giá trình độ chặt chẽ của mối liên hệ tương quan.

$$\eta = \sqrt{\frac{\delta_{y_x}^2}{\delta_y^2}} = \sqrt{1 - \frac{\delta_{y_{(x)}}^2}{\delta_y^2}}$$

Hay:

$$\eta = \sqrt{1 - \frac{\sum (y_i - y_x)^2}{\sum (y_i - \bar{y})^2}}$$

Trong đó: η - Tỷ số tương quan ;

y_i - Là trị số thực nghiệm của tiêu thức kết quả;

y_x - Là trị số lý thuyết của tiêu thức kết quả, tính theo phương trình hồi qui;

$\bar{y}$: Là số trung bình số học của các trị số thực nghiệm.

Tính chất của tỷ số tương quan:

Tỷ số tương quan lấy giá trị trong khoảng $0 \leq \eta \leq 1$

+ Nếu $\eta = 0$ giữa x và y không có mối liên hệ tương quan;

+ Nếu $\eta = 1$ giữa x và y có liên hệ hàm số;

+ Nếu η càng gần 1 mối liên hệ giữa x và y càng chặt chẽ và ngược lại.

Trong cùng một tiêu thức nghiên cứu x và y tỷ số tương quan bao giờ cũng lớn hơn giá trị tuyệt đối của hệ số tương quan tuyến tính. Nếu $\eta > |r|$ thì giữa x và y có mối liên hệ tương quan phi tuyến. Còn lại nếu $\eta = |r|$ thì giữa x và y có mối liên hệ tương quan tuyến tính.

5.4. LIÊN HỆ TƯƠNG QUAN TUYẾN TÍNH GIỮA NHIỀU TIÊU THỨC SỐ LƯỢNG

Trong thực tế một tiêu thức thường có nhiều mối liên hệ với các tiêu thức khác. Khi phân tích mối liên hệ giữa nhiều tiêu thức,

cần căn cứ vào mục đích nghiên cứu để chọn các tiêu thức có liên hệ với nhau và chỉ chọn những tiêu thức có ý nghĩa nhất, ảnh hưởng lớn nhất đối với tiêu thức kết quả.

Để biểu diễn mối liên hệ tương quan giữa nhiều tiêu thức số lượng thường dùng phương trình tuyến tính. Phương trình tổng quát có dạng:

$$\bar{y}_{x_1 x_2 \dots x_n} = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_n x_n$$

Trong đó: $x_1, x_2, \dots, x_n$ - Là tiêu thức nguyên nhân;

y - Là tiêu thức kết quả;

$a_0, a_1, a_2, \dots, a_n$ - Là các tham số của phương trình hồi qui, thường được xác định bằng phương pháp bình phương nhỏ nhất.

Để đánh giá trình độ chặt chẽ của mối liên hệ tương quan tuyến tính giữa nhiều tiêu thức số lượng người ta thường sử dụng hệ số tương quan bội.

$$R_{\bar{y}_{x_1 x_2 \dots x_n}} = \sqrt{1 - \frac{\sum (y - \bar{y}_{x_1 x_2 \dots x_n})^2}{\sum (y - \bar{y})^2}}$$

Trong đó: $R_{\bar{y}_{x_1 x_2 \dots x_n}}$ - Hệ số tương quan bội, hệ số tương quan bội lấy giá trị trong khoảng $[0, 1]$. Nếu $R = 1$ thì có mối liên hệ hàm số;

$\bar{y}_{x_1 x_2 \dots x_n}$ - Giá trị lý thuyết tính theo phương trình hồi quy;

y - Giá trị thực nghiệm;

$\bar{y}$ - Số trung bình số học của các giá trị thực nghiệm.

Ví dụ: Giả sử có phương trình biểu diễn mối liên hệ tương quan tuyến tính giữa 03 tiêu thức, trong đó có 02 tiêu thức nguyên nhân (x_1 và x_2) và tiêu thức kết quả (y):

$$\bar{y}_{x_1x_2} = a_0 + a_1x_1 + a_2x_2$$

Các tham số a_0, a_1, a_2 được xác định bằng phương pháp bình phương nhỏ nhất bằng cách giải hệ phương trình chuẩn tắc:

$$\begin{cases} na_0 + a_1 \sum x_1 + a_2 \sum x_2 = \sum y \\ a_0 \sum x_1 + a_1 \sum x_1^2 + a_2 \sum x_1x_2 = \sum x_1y \\ a_0 \sum x_2 + a_1 \sum x_1x_2 + a_2 \sum x_2^2 = \sum yx_2 \end{cases}$$

Để đánh giá trình độ chặt chẽ mối liên hệ tương quan sử dụng hệ số tương quan bội theo công thức:

$$R_{\bar{y}_{x_1x_2}} = \sqrt{\frac{r_{yx_1}^2 + r_{yx_2}^2 - 2r_{yx_1} r_{yx_2} r_{x_1x_2}}{1 - r_{x_1x_2}^2}}$$

Trong đó

$$r_{yx_1} = \frac{\overline{x_1y} - \bar{x}_1\bar{y}}{\delta_{x_1}\delta_y}; \quad r_{yx_2} = \frac{\overline{x_2y} - \bar{x}_2\bar{y}}{\delta_{x_2}\delta_y}; \quad r_{x_1x_2} = \frac{\overline{x_1x_2} - \bar{x}_1\bar{x}_2}{\delta_{x_1}\delta_{x_2}}$$

Để tiến hành phân tích sâu sắc hơn, đánh giá trình độ chặt chẽ mối liên hệ tương quan riêng giữa từng tiêu thức nguyên nhân x_1, x_2 với tiêu thức kết quả y dùng hệ số tương quan riêng.

Hệ số tương quan giữa tiêu thức x_1 và tiêu thức y :

$$r_{yx_1(x_2)} = \frac{r_{yx_1}^2 - r_{yx_2} r_{x_1x_2}}{\sqrt{(1 - r_{yx_2}^2)(1 - r_{x_1x_2}^2)}}$$

Chương 6

DÃY SỐ THỜI GIAN

6.1. KHÁI NIỆM, PHÂN LOẠI VÀ Ý NGHĨA CỦA DÃY SỐ THỜI GIAN

6.1.1. Khái niệm

Một lượng của các hiện tượng kinh tế không ngừng biến động theo thời gian. Để nghiên cứu sự biến động này, người ta thường sử dụng dãy số thời gian. Dãy số thời gian là dãy các trị số của chỉ tiêu thống kê được sắp xếp theo thứ tự thời gian.

Ví dụ: Số lượng phút đàm thoại đường dài của một Bưu cục qua các năm như sau:

Bảng 6.1

Năm	1998	1999	2000	2001	2002	2003	2004
Lưu lượng (10 ³ phút)	100	110	108	126	140	141	138

Thành phần dãy số thời gian: Mỗi dãy số thời gian có hai thành phần: thời gian và mức độ của chỉ tiêu về hiện tượng nghiên cứu, trong đó:

- *Thời gian của dãy số:* Tùy theo mục đích nghiên cứu và chỉ tiêu biến đổi nhiều hay ít mà có thể biểu thị bằng ngày, tháng, năm. Độ dài giữa hai khoảng thời gian kế tiếp nhau gọi là khoảng thời gian; có trường hợp khoảng thời gian đều nhau, có trường hợp khoảng thời gian là không đều nhau.

- *Mức độ của chỉ tiêu về hiện tượng*: Được phản ánh bằng các trị số của chỉ tiêu gọi là các mức độ của dãy số và nó có thể biểu diễn bằng số tuyệt đối, số tương đối hay số trung bình.

6.1.2. Phân loại dãy số thời gian

Căn cứ vào đặc điểm của yếu tố thời gian trong dãy số thì dãy số thời gian được chia thành 02 loại:

- *Dãy số thời kỳ*: Là dãy số phản ánh mặt lượng của hiện tượng qua từng thời kỳ, tức là dãy số phản ánh quy mô, khối lượng của hiện tượng nghiên cứu trong từng khoảng thời gian nhất định (năm, tháng, quý). Khoảng thời gian trong dãy số càng dài thì trị số của chỉ tiêu càng lớn vì thế có thể cộng các trị này với nhau để phản ánh mặt lượng của hiện tượng trong thời kỳ dài hơn.

- *Dãy số thời điểm*: Phản ánh mặt lượng của hiện tượng nghiên cứu tại các thời điểm nhất định. Mức độ của thời điểm sau thường bao gồm toàn bộ hoặc một bộ phận mức độ của thời điểm trước đó. Vì vậy, việc cộng các trị số của chỉ tiêu không phản ánh quy mô của hiện tượng.

6.1.3. Ý nghĩa của dãy số thời gian

Dãy số thời gian giúp cho thống kê nghiên cứu các đặc điểm về sự biến động của hiện tượng và tính qui luật của phát triển của hiện tượng theo thời gian, đồng thời để dự đoán mức độ của hiện tượng trong tương lai.

6.1.4. Yêu cầu đối với dãy số thời gian

Để phản ánh một cách chính xác sự phát triển của hiện tượng theo thời gian thì khi xây dựng một dãy số thời gian phải đảm bảo

tính chất có thể so sánh được giữa các mức độ trong dãy số. Cụ thể là nội dung và phương pháp tính chỉ tiêu qua các thời gian khác nhau phải thống nhất. Phạm vi của tổng thể nghiên cứu trước sau phải nhất quán. Các khoảng cách thời gian trong dãy số nên bằng nhau để tiện lợi cho việc tính toán.

6.2. CÁC CHỈ TIÊU PHÂN TÍCH DÃY SỐ THỜI GIAN

6.2.1. Mức độ bình quân theo thời gian

Mức độ bình quân theo thời gian là chỉ tiêu tổng hợp phản ánh mức độ điển hình của hiện tượng trong toàn bộ khoảng thời gian nghiên cứu, hoặc từng giai đoạn nghiên cứu.

Ký hiệu: $y_1, y_2, \dots, y_n$ - Là các mức độ của dãy số thời gian;

$\bar{y}$ - Mức độ trung bình theo thời gian.

Tùy theo dãy số thời kỳ hay thời điểm mà có các công thức tính khác nhau:

- Đối với dãy số thời kỳ

$$\bar{y} = \frac{y_1 + y_2 + y_3 + \dots + y_n}{n} = \frac{\sum_{i=1}^n y_i}{n}$$

Trong đó : y_i ($i = 1, 2, \dots, n$) - Là mức độ thứ i trong dãy số;

n - Là số các mức độ của dãy số.

- Đối với dãy số thời điểm

Dãy số thời điểm phản ánh qui mô, khối lượng của hiện tượng trong từng thời điểm. Nếu muốn tính mức độ trung bình chính xác người ta phải xác định trị số chỉ tiêu từng ngày. Nhưng trên thực tế, chỉ có trị số chỉ tiêu vào một ngày nào đó trong

tháng nên phải giả thiết rằng người giữa hai thời điểm điều tra mật độ của hiện tượng tăng giảm đều đặn.

+ Dãy số thời điểm có khoảng cách thời gian bằng nhau

$$\bar{y} = \frac{\frac{y_1 + y_2}{2} + \frac{y_2 + y_3}{2} + \frac{y_3 + y_4}{2} + \dots + \frac{y_{n-1} + y_n}{2}}{n-1}$$

$$\Rightarrow \bar{y} = \frac{\frac{y_1}{2} + y_2 + y_3 + \dots + \frac{y_n}{2}}{n-1}$$

Trong đó: y_i - Mức độ thứ i của dãy số thời điểm có khoảng cách thời gian bằng nhau.

+ Dãy số thời điểm có khoảng cách thời gian không bằng nhau

$$\bar{y} = \frac{\sum_{i=1}^n y_i t_i}{\sum_{i=1}^n t_i}$$

Trong đó: y - Là mức độ thứ i trong dãy số;

t_i - Là độ dài thời gian có mức độ y_i .

Ví dụ: Có tài liệu về số công nhân có trong danh sách của một đơn vị Bưu điện trong quý I năm 2004 như sau:

Ngày 01/01 có 50 công nhân

Ngày 04/02 thêm 3 công nhân

Ngày 15/3 giảm 6 công nhân

Ngày 30/3 thêm 5 công nhân

Tính số công nhân bình quân của đơn vị Bưu điện đó trong tháng của quý I năm 2002. Để xác định số công nhân bình quân trong tháng của đơn vị lập bảng:

Bảng 6.2

Thời gian	Số ngày t_i	Số công nhân y_i
Từ 01/01 đến 03/02	34	50
Từ 04/02 đến 14/3	39	53
Từ 15/3 đến 29/3	45	47
Từ 30/3 đến 31/3	2	52

Do đó:

$$\bar{y} = \frac{50.34 + 53.39 + 47.15 + 52.2}{34 + 39 + 15 + 2} = 51 \text{ (công nhân)}$$

6.2.2. Lượng tăng (hoặc giảm) tuyệt đối

Là chỉ tiêu phản ánh sự thay đổi tuyệt đối về mức độ của hiện tượng theo thời gian. Nó được xác định bằng hiệu số giữa hai mức độ trong dãy số. Tùy theo mục đích nghiên cứu mà có các cách tính:

6.2.2.1. Lượng tăng (hoặc giảm) tuyệt đối liên hoàn (từng kỳ)

Là chênh lệch giữa mức độ kỳ nghiên cứu (y_i) với mức độ của kỳ đứng liền trước nó (y_{i-1}), nhằm phản ánh mức độ tăng (hoặc giảm) tuyệt đối giữa hai thời gian liền nhau:

$$\Delta_i = y_i - y_{i-1} \quad (i = 2 \dots n)$$

6.2.2.2. Lượng tăng (hoặc giảm) tuyệt đối định gốc

Là hiệu số giữa mức độ của kỳ nghiên cứu (y_i) với mức độ của một kỳ được chọn làm gốc cố định - thường là mức độ đầu tiên của dãy số y_1 , nhằm để phản ánh mức độ tăng (giảm) của hiện tượng trong khoảng thời gian dài.

$$\Delta_i^* = y_i - y_1 \quad (i = 2 \dots n)$$

Xét các trường hợp sau:

- Nếu $\Delta_i, \Delta_i^* > 0$ thì mức độ hiện tượng tăng;
- Nếu $\Delta_i, \Delta_i^* = 0$ thì mức độ hiện tượng không thay đổi;
- Nếu $\Delta_i, \Delta_i^* < 0$ thì mức độ hiện tượng giảm;
- Tổng đại số của các lượng tăng (giảm) tuyệt đối từng kỳ bằng lượng tăng (giảm) tuyệt đối định gốc tương ứng.

$$(y_2 - y_1) + (y_3 - y_2) + \dots + (y_i - y_{i-1}) = y_i - y_1 \Leftrightarrow \sum \Delta_i = \Delta_i^*$$

6.2.2.3. Lượng tăng (hoặc giảm) tuyệt đối bình quân

Là số trung bình cộng của các lượng tăng (hoặc giảm) tuyệt đối liên hoàn. Nó phản ánh mức độ tăng trung bình của hiện tượng nghiên cứu trong thời kỳ dài:

$$\overline{\Delta y} = \sum_2^n \Delta_i = \frac{(y_2 - y_1) + (y_3 - y_2) + \dots + (y_i - y_{i-1})}{n - 1} = \frac{y_n - y_1}{n - 1}$$

Trong đó: $\overline{\Delta y}$ - nói lên trong một đơn vị thời gian mức độ của hiện tượng tăng (hoặc giảm) bao nhiêu.

6.2.3. Tốc độ phát triển

Là chỉ tiêu tương đối phản ánh xu hướng phát triển của hiện tượng nghiên cứu qua thời gian. Chỉ tiêu này được xác định bằng tỷ số giữa hai mức độ của hiện tượng ở hai thời kỳ hoặc hai thời điểm.

Tùy theo mục đích nghiên cứu, ta có các loại tốc độ phát triển sau đây:

6.2.3.1. Tốc độ phát triển liên hoàn (từng kỳ)

Là tỷ số giữa mức độ của kỳ nghiên cứu (y_i) với mức độ của kỳ đứng ngay trước đó (y_{i-1}). Chỉ tiêu này phản ánh sự phát triển của hiện tượng giữa hai thời gian liên nhau:

$$t_i = \frac{y_i}{y_{i-1}} \quad (i = 2, 3, \dots, n)$$

hay
$$t_i = \frac{y_i}{y_{i-1}} \times 100 (\%)$$

6.2.3.2. Tốc độ phát triển định gốc

Là tỷ số giữa mức độ của kỳ nghiên cứu (y_i) với mức độ của một kỳ được chọn làm gốc cố định, thường là mức độ đầu tiên của dãy số (y_1). Chỉ tiêu này biểu hiện sự phát triển của hiện tượng trong các khoảng thời gian dài.

$$T_i = \frac{y_i}{y_1} \quad \text{hay} \quad T_i = \frac{y_i}{y_1} \times 100 \quad (i = 2, 3, \dots, n)$$

Giữa tốc độ phát triển liên hoàn và tốc độ phát triển định gốc có mối quan hệ sau:

- Tích các tốc độ phát triển liên hoàn bằng tốc độ phát triển định gốc, nghĩa là:

$$\frac{y_2}{y_1} \cdot \frac{y_3}{y_2} \cdot \frac{y_4}{y_3} \dots \frac{y_i}{y_{i-1}} = \frac{y_i}{y_1} \Leftrightarrow \prod t_i = T_i$$

- Thương của hai tốc độ phát triển định gốc liên nhau bằng tốc độ phát triển liên hoàn giữa hai thời gian:

$$\frac{T_i}{T_{i-1}} = \frac{\frac{y_i}{y_1}}{\frac{y_{i-1}}{y_1}} = \frac{y_i}{y_{i-1}} = t_i \quad (i = 2, 3, \dots, n)$$

6.2.3.3. Tốc độ phát triển bình quân

Là số trung bình cộng của các tốc độ phát triển liên hoàn. Chỉ tiêu này biểu hiện tốc độ phát triển trung bình của hiện tượng trong suốt thời gian nghiên cứu.

$$\bar{t} = \sqrt[n]{t_1 t_2 \dots t_n} = \sqrt[n]{\prod_{i=1}^n t_i} = \sqrt[n]{\frac{y_n}{y_1}}$$

6.2.4. Tốc độ tăng (hoặc giảm)

Là chỉ tiêu tương đối phản ánh mức độ của hiện tượng nghiên cứu giữa 2 thời gian đã tăng hoặc giảm bao nhiêu lần (hoặc bao nhiêu phần trăm). Tùy theo mục đích nghiên cứu ta có:

6.2.4.1. Tốc độ tăng (hoặc giảm) liên hoàn

Là tỷ số giữa lượng tăng (hoặc giảm) tuyệt đối liên hoàn với mức độ kỳ gốc liên hoàn (kỳ gốc đứng ngay trước nó).

$$a_i = \frac{y_i - y_{i-1}}{y_{i-1}} = t_i - 1 \quad (i = 2, 3, \dots, n)$$

Hay $a_i = t_i - 100 (\%)$

6.2.4.2. Tốc độ tăng (hoặc giảm) định gốc

Là tỷ số giữa lượng tăng (hoặc giảm) tuyệt đối định gốc và mức độ kỳ gốc cố định:

$$A_i = \frac{y_i - y_1}{y_1} = t_i - 1 \quad (i = 2, 3, \dots, n)$$

Hay $A_i = T_i - 100 (\%)$

Vậy tốc độ tăng (giảm) liên hoàn hay định gốc bằng tốc độ phát triển liên hoàn hay định gốc trừ 1 (hay 100%).

6.2.4.3. Tốc độ tăng (hoặc giảm) bình quân

Là chỉ tiêu tương đối nói lên nhịp điệu tăng (hoặc giảm) điển hình của hiện tượng nghiên cứu trong thời gian dài.

$$\bar{a} = \bar{t} - 1$$

Nếu $\bar{t}$ tính bằng % thì $\bar{a} = \bar{t} - 100$.

6.2.5. Giá trị tuyệt đối của 1% tăng (hoặc giảm)

Chỉ tiêu này phản ánh cứ 1% tăng (hoặc giảm) của tốc độ tăng (hoặc giảm) liên hoàn thì tương ứng với một trị số tuyệt đối là bao nhiêu?

$$g_i = \frac{\Delta_i}{a_i} = \frac{y_i - y_{i-1}}{\frac{y_i - y_{i-1}}{y_{i-1}}} = \frac{y_{i-1}}{100} \quad (i = 2, 3, \dots, n)$$

Với a_i tính bằng %.

6.3. CÁC PHƯƠNG PHÁP BIỂU HIỆN XU HƯỚNG PHÁT TRIỂN CỦA HIỆN TƯỢNG

Trong khi phân tích các dãy số thời gian yêu cầu phải thể hiện rõ được xu hướng phát triển cơ bản của hiện tượng. Trong thực tế biến động của hiện tượng chịu ảnh hưởng của nhiều nhân tố. Ngoài các nhân tố chủ yếu, cơ bản quyết định xu hướng phát triển của hiện tượng, còn có những nhân tố ngẫu nhiên gây ra những sai lệch khỏi xu hướng. Vì vậy cần phải sử dụng các phương pháp thích hợp để loại trừ ảnh hưởng của các nhân tố ngẫu nhiên đó nhằm nêu rõ xu hướng và tính quy luật của sự phát triển. Người ta có các phương pháp thường dùng sau:

6.3.1. Phương pháp mở rộng khoảng cách thời gian

Phương pháp này được áp dụng khi dãy số có khoảng cách thời gian tương đối ngắn và có quá nhiều mức độ nên không phản ánh được xu hướng phát triển của hiện tượng. Người ta có thể rút bớt số các mức độ trong dãy số bằng phương pháp mở rộng khoảng cách thời gian để có dãy số mới có các mức độ trong đó sự tác động của các nhân tố ngẫu nhiên bị loại trừ (bởi chúng tác động theo chiều hướng trái ngược nhau)

Ví dụ: Có số liệu về doanh thu kinh doanh năm 2004 của đơn vị Bưu điện Y như sau:

Bảng 6.3

Tháng	Doanh thu kinh doanh (10 ⁶ đồng)	Tháng	Doanh thu kinh doanh (10 ⁶ đồng)
1	40,5	7	40,6
2	36,7	8	45,0
3	40,4	9	49,2
4	38,2	10	49,1
5	42,0	11	46,0
6	48,7	12	42,4

Dãy số trên cho thấy doanh thu kinh doanh theo các tháng khi tăng, khi giảm thất thường. Do đó có thể mở rộng khoảng cách thời gian từ tháng sang quý:

Bảng 6.4

Quý	Doanh thu kinh doanh (10^6 đồng)
I	117,6
II	128,9
III	134,8
IV	137,5

Do khoảng cách thời gian được mở rộng từ tháng sang quý, nên mỗi mức độ của dãy số mới tác động của các nhân tố ngẫu nhiên được loại trừ và do đó ta thấy tình hình thực hiện doanh thu kinh doanh tăng từ quý I đến quý IV.

6.3.2. Phương pháp số bình quân trượt (di động)

Phương pháp này dùng để điều chỉnh các mức độ trong dãy số có biến động tăng giảm thất thường nhằm loại trừ ảnh hưởng của các nhân tố ngẫu nhiên, vạch rõ xu hướng phát triển cơ bản của hiện tượng.

Số bình quân di động (trượt) là số trung bình cộng của một nhóm nhất định các mức độ của dãy số được tính bằng cách lần lượt loại trừ dần các mức độ đầu, đồng thời thêm vào các mức độ tiếp theo, sao cho số lượng các mức độ tham gia tính số trung bình không thay đổi.

Giả sử có dãy số thời gian:

Bảng 6.5

Thời gian	t_1	t_2	t_3	...	t_n
Mức độ y	y_1	y_2	y_3	...	y_n

Nếu tính số trung bình trượt từ nhóm $m = 3$ mức độ. Ta có:

$$\bar{y}_2 = \frac{y_1 + y_2 + y_3}{3}$$

$$\bar{y}_3 = \frac{y_2 + y_3 + y_4}{3}$$

...

$$\bar{y}_{n-1} = \frac{y_{n-2} + y_{n-1} + y_n}{3}$$

Từ đó có thể xây dựng dãy số thời gian gồm các số trung bình trượt:

$$\bar{y}_2, \bar{y}_3, \dots, \bar{y}_{n-1}$$

Vấn đề được đặt ra là nên tính số trung bình trượt từ một nhóm bao nhiêu mức độ?

Nếu tính số trung bình trượt từ một nhóm ít mức độ thì ảnh hưởng của các nhân tố ngẫu nhiên ít bị loại trừ. Tuy nhiên ta sẽ có nhiều số trung bình trượt và do đó sẽ dễ dàng đánh giá xu hướng biến động của hiện tượng. Ngược lại nếu số trung bình trượt được tính từ một nhóm nhiều mức độ thì khả năng hạn chế, loại bỏ ảnh hưởng ngẫu nhiên sẽ lớn. Tuy nhiên số lượng số trung bình trượt tính được sẽ ít hơn, và có thể gây khó khăn trong việc đánh giá xu hướng phát triển của hiện tượng.

Do vậy trên thực tế nghiên cứu người ta thường làm như sau:

- Đối với những hiện tượng biến động không lớn và số mức độ thực tế không nhiều lắm thì số trung bình trượt có thể tính từ một nhóm 03 mức độ. Nếu biến động của hiện tượng lớn thì nên tính số trung bình trượt từ một nhóm nhiều mức độ hơn (5, 7, ... mức độ).

- Nếu hiện tượng biến động theo chu kỳ thì nên chọn thời kỳ tính số trung bình di động bằng với độ dài thời gian (hoặc bội số) của chu kỳ. Ví dụ nếu chu kỳ biến động là 3, 5 năm thì có thể tính số trung bình di động với 7 mức độ, hoặc với dãy số biến động thời vụ theo tháng, quý thì có thể tính số trung bình di động từ nhóm 12 hoặc 4 mức độ.

6.3.3. Phương pháp hồi qui

6.3.3.1. Thực chất của phương pháp hồi qui

Là căn cứ vào đặc điểm biến động của các mức độ trong dãy số thời gian người ta tìm một hàm số (gọi là phương trình hồi qui) nhằm phản ánh sự biến động của hiện tượng theo thời gian.

Việc lựa chọn dạng của phương trình hồi qui (đường thẳng hay đường cong) phụ thuộc vào số liệu thống kê thực tế và phân tích đặc điểm biến động của hiện tượng qua thời gian, đồng thời kết hợp với một số phương pháp đơn giản khác.

Với biến thời gian t dạng tổng quát của phương trình hồi qui được biểu diễn như sau: $\bar{y}_t = f(t, a_0, a_1, \dots, a_n)$

Trong đó: $\bar{y}_t$ - Là các giá trị lý thuyết;

t - Là thứ tự thời gian;

$a_0, a_1, a_2, \dots, a_n$ - Là các tham số, được xác định bằng phương pháp tổng bình phương nhỏ nhất. Tức là:

$$S = \sum_{i=1}^n (y_i - \bar{y}_{t_i})^2 \Rightarrow \min$$

Trong đó: y_i - Là mức độ thực tế thứ i ;

$\bar{y}_{t_i}$ - Là mức độ lý thuyết thứ i .

6.3.3.2. Các dạng phương trình hồi qui

a) Phương trình đường thẳng

$$\bar{y}_t = a_0 + a_1 t$$

Các tham số a_0, a_1 được xác định bằng phương pháp bình phương nhỏ nhất.

$$S = \sum (y - a_0 - a_1 t)^2 \rightarrow \min \rightarrow$$

$$\rightarrow \begin{cases} \frac{\partial S}{\partial a_0} = 0 \rightarrow 2 \sum (y - a_0 - a_1 t)(-1) = 0 \\ \frac{\partial S}{\partial a_1} = 0 \rightarrow 2 \sum (y - a_0 - a_1 t)(-t) = 0 \end{cases}$$

$$\rightarrow \begin{cases} \sum y - n a_0 - a_1 \sum t = 0 \\ \sum y t - a_0 \sum t - a_1 \sum t^2 = 0 \end{cases} \rightarrow \begin{cases} n a_0 + a_1 \sum t = \sum y \\ a_0 \sum t + a_1 \sum t^2 = \sum y t \end{cases}$$

Đây là hệ phương trình chuẩn tắc để tìm các tham số của phương trình.

Trong trường hợp này ta thấy t là thứ tự thời gian, cho nên có thể thay thế t bằng t' với $\sum t' = 0$, sao cho vẫn đảm bảo tính thứ tự, lúc đó việc tính toán sẽ đơn giản hơn. Khi sử dụng phép biến đổi cần lưu ý:

- Nếu số lượng thời gian là một số lẻ thì lấy thời gian đứng giữa bằng 0, các thời gian đứng trước lần lượt là -1, -2,... và thời gian đứng sau lần lượt là +1, +2 ...

- Nếu số lượng thời gian là một số chẵn thì lấy hai thời gian đứng giữa là -1 và +1, các thời gian đứng trước lần lượt là -3, -5,... và thời gian đứng sau lần lượt là +3, +5,...

Với $\Sigma t' = 0$ hệ phương trình trên sẽ có dạng:

$$\begin{cases} na_0 = \Sigma y \\ a_1 \Sigma t'^2 = \Sigma yt' \end{cases}$$

Giải hệ phương trình này ta sẽ xác định được các tham số a_0 , a_1 .

Ví dụ: Có số liệu doanh thu của một đơn vị Bưu điện như sau:

Bảng 6.6

Năm	Doanh thu (10 ⁹ đồng)	Phần tính toán			
		Thứ tự thời gian	t^2	$t.y$	y_t
1999	91,0	1	1	91,0	91,93
2000	96,9	2	4	192,4	96,68
2001	102,2	3	9	306,6	101,43
2002	106,5	4	16	426,0	106,18
2003	110,3	5	25	551,5	110,93
2004	115,9	6	36	695,4	115,68
Cộng	622,8	21	91	2262,9	

Theo hệ phương trình chuẩn tắc, ta có:

$$\begin{cases} 6a_0 + 21a_1 = 622,8 \\ 21a_0 + 91a_1 = 2262,9 \end{cases}$$

Giải hệ phương trình ta được $a_0 = 87,18$; $a_1 = 4,75$. Phương trình hồi qui $\bar{y}_t = 87,18 + 4,75.t$; $a_1 = 4,75$ phản ánh mức tăng bình quân hàng năm. Thay các giá trị của $t = 1, 2, \dots$ vào phương trình, sẽ được các giá trị lý thuyết $\bar{y}_t$.

Thay $t = t'$ sao cho $\sum t' = 0$

Bảng 6.7

Năm	y	Phần tính toán			
		t'	t' ²	t'.y	$\bar{y}_t$
1999	91,0	-5	25	-455,0	91,75
2000	96,9	-3	9	-290,7	96,57
2001	102,2	-1	1	-102,2	101,39
2002	106,5	1	1	106,5	106,21
2003	110,3	3	9	330,9	111,03
2004	115,9	5	25	579,5	115,85
Cộng	622,8	0	70	169,0	

$$\begin{cases} 6a'_0 = 622,8 \\ 70a'_1 = 169 \end{cases} \Rightarrow \begin{cases} a'_0 = \frac{622,8}{6} = 103,8 \\ a'_1 = \frac{169}{70} = 2,41 \end{cases}$$

Do đó $\bar{y}_t = 103,8 + 2,41.t'$. Thay các giá trị của t' vào phương trình trên sẽ có các giá trị $\bar{y}_t$.

Tác dụng của phương pháp hồi qui: Ngoài việc biểu diễn xu hướng của hiện tượng còn có khả năng dự báo hiện tượng trong tương lai. Muốn có dự đoán hiện tượng vào thời điểm nào ta thay t vào phương trình hồi qui.

b) Phương trình Hypebol

Phương trình hồi qui có dạng: $\bar{y}_t = a_0 + \frac{a_1}{t}$

Các tham số a_0, a_1 được xác định bởi hệ phương trình:

$$\begin{cases} na_0 + a_1 \sum \frac{1}{t} = \sum y \\ a_0 \sum \frac{1}{t} + a_1 \sum \frac{1}{t^2} = \sum \frac{1}{t} y \end{cases}$$

c) Phương trình hàm mũ

Phương trình hồi qui có dạng:

$$\bar{y}_t = a_0 a_1^t$$

Lô-ga-rít hóa hai vế của phương trình ta được:

$$\lg y_t = \lg a_0 + t \lg a_1$$

Các tham số của phương trình được xác định từ hệ phương trình chuẩn tắc:

$$\begin{cases} n \lg a_0 + \lg a_1 \sum t = \sum \lg y_t \\ \lg a_0 \sum t + \lg a_1 \sum t^2 = \sum t \lg y_t \end{cases}$$

d) Phương trình Parabol bậc 2

Phương trình hồi qui có dạng:

$$\bar{y}_t = a_0 + a_1 t + a_2 t^2$$

Các tham số a_0 , a_1 , a_2 được xác định bằng cách giải hệ phương trình chuẩn tắc:

$$\begin{cases} na_0 + a_1 \sum t + a_2 \sum t^2 = \sum y \\ a_0 \sum t + a_1 \sum t^2 + a_2 \sum t^3 = \sum yt \\ a_0 \sum t^2 + a_1 \sum t^3 + a_2 \sum t^4 = \sum yt^2 \end{cases}$$

6.3.4. Phương pháp biểu hiện biến động thời vụ

Sự biến động của một số hiện tượng kinh tế xã hội thường mang tính thời vụ, nghĩa là hàng năm, trong từng thời gian nhất định, sự biến động được lặp đi, lặp lại.

Biến động thời vụ thường gây ra tình trạng làm ảnh hưởng đến hoạt động sản xuất kinh doanh của ngành và các ngành có liên quan. Ví dụ như sự thay đổi nhu cầu đi lại của dân chúng theo thời gian trong năm (tăng lên nhiều vào các dịp nghỉ hè, tết, ngày lễ...) sẽ gây khó khăn cho ngành vận tải trong việc bố trí phương tiện và nhân lực để phục vụ hết nhu cầu của nhân dân. Nghiên cứu biến động thời vụ cho phép chủ động trong công tác quản lý kinh tế - xã hội, hạn chế những ảnh hưởng của biến động thời vụ đối với sản xuất kinh doanh và đời sống xã hội.

Để xác định tính chất và mức độ biến động thời vụ người ta thường dùng phương pháp tính các chỉ số thời vụ dựa trên số liệu của nhiều năm (ít nhất là 03 năm).

Đối với dãy số thời gian tương đối ổn định, tức là dãy số trong đó mức độ của hiện tượng từ năm này qua năm khác không có biểu hiện tăng giảm rõ rệt:

$$I_{tvi} = \frac{\bar{y}_i}{\bar{y}_0} \times 100$$

Trong đó: $\bar{y}_i$ - Số trung bình của tất các các mức độ các tháng cùng tên i;

$\bar{y}_0$ - Số trung bình của tất cả các mức độ trong dãy số;

I_{tvi} - Chỉ số thời vụ.

Ví dụ: Khối lượng sản phẩm EMS của một đơn vị Bưu điện X trong 03 năm như sau:

Bảng 6.8

Tháng	Khối lượng sản phẩm EMS			$\bar{y}_i$	$I_{tvi} = \frac{\bar{y}_i}{\bar{y}_0} \times 100$
	2002	2003	2004		
1	1.495	1.500	1.490	1.495	62,9
2	1.461	1.490	1.480	1.477	62,2
3	1.533	1.599	1.604	1.578	66,4
4	1.922	2.210	2.005	2.046	86,1
5	2.476	2.804	2.745	2.765	116,4
6	3.289	3.282	3.250	3.274	137,8
7	3.523	3.620	3.700	3.614	152,1
8	3.330	3.300	3.215	3.282	138,2
9	2.597	2.604	2.599	2.957	109,3
10	2.249	2.205	2.304	2.253	94,8
11	2.144	2.200	2.190	2.178	91,7
12	1.983	1.889	1.950	1.941	81,7
Cộng	28.272	28.703	28.523	$\bar{y}_i = 2.375$	-

Từ kết quả tính các chỉ số thời vụ cho thấy khối lượng sản phẩm EMS nhiều từ tháng 5 đến tháng 9 (cao nhất là tháng 7), sau đó giảm dần và mức thấp nhất là tháng 02.

Trong trường hợp các mức độ của dãy số năm này qua năm khác biến động nhiều, thì chỉ số thời vụ được tính theo công thức sau:

$$\bar{I}_{vi} = \frac{\sum \frac{y_i}{\bar{y}_{li}}}{n} \times 100$$

Trong đó: y_i - Mức độ thực tế thứ i ;

$\bar{y}_{li}$ - Mức độ lý thuyết thứ i tính theo phương trình hồi qui hoặc phương pháp số trung bình di động.

6.4. DỰ BÁO THỐNG KÊ NGẮN HẠN

6.4.1. Khái niệm và đặc điểm của dự báo thống kê

Dự báo thống kê là xác định các mức độ có thể xảy ra trong tương lai của hiện tượng nghiên cứu. Biết được tương lai của hiện tượng sẽ giúp các nhà quản trị chủ động cũng như có những quyết định đúng trong kinh doanh.

Xuất phát từ đối tượng và nhiệm vụ nghiên cứu, từ nguồn tài liệu thống kê thích hợp, thống kê thường thực hiện dự đoán ngắn hạn gọi là dự đoán thống kê ngắn hạn.

Dự báo thống kê ngắn hạn là công cụ quan trọng để tổ chức quản lý một cách thường xuyên các hoạt động sản xuất kinh doanh từ ngành đến các cấp cơ sở, nó cho phép phát hiện những nhân tố mới, những sự mất cân đối để từ đó có biện pháp phù hợp trong quá trình quản lý.

Có nhiều phương pháp dự báo khác nhau, phụ thuộc vào nguồn thông tin cũng như mục tiêu của dự đoán. Nhưng nội dung cơ bản của dự báo thống kê là dựa trên các giá trị đã biết $y_1, y_2, \dots, y_n$, phân tích các yếu tố ảnh hưởng đến sự biến động của hiện tượng, thừa nhận rằng những yếu tố đã và đang tác động sẽ vẫn còn tác động đến hiện tượng trong tương lai, xây dựng mô hình để dự đoán các giá trị tương lai chưa biết của hiện tượng.

Dự báo thống kê có những đặc điểm cơ bản sau đây:

- Dự báo thống kê chỉ thực hiện được trên từng mô hình cụ thể. Tức là nó chỉ thực hiện được sau khi đã phân tích thực trạng biến động theo thời gian hoặc không gian và phân tích đánh giá các nguyên nhân ảnh hưởng đến tiêu thức kết quả. Trong phân tích thống kê cần phân biệt rõ 02 mô hình cơ bản sau:

+ Mô hình dãy số thời gian: Là tính quy luật biến động của hiện tượng qua thời gian được biểu hiện bằng hàm xu thế trên cơ sở phân tích sự biến động dãy số tiền sử trong quá khứ, hiện tại và tiến tới tương lai.

+ Mô hình nhân quả: Là mối quan hệ nhân quả giữa các hiện tượng nghiên cứu qua thời gian hoặc không gian được biểu hiện bằng các hàm kinh tế, phương trình kinh tế, phương trình tương quan.

Do đó, dự báo thống kê không phải là sự phán đoán theo định tính hoặc “đoán mò” mà là sự định lượng điều sẽ xảy ra, khả năng sẽ xảy ra nhiều nhất hoặc định lượng mức độ phải xảy ra trên cơ sở khoa học của phân tích thực tiễn, cho nên kết quả dự báo thống kê vừa mang tính khách quan vừa mang tính chủ quan và nó phụ thuộc vào trình độ nhận thức khách quan, hay khả năng tư duy của người dự báo.

- Nguyên tắc cơ bản để xác định mô hình dự báo là tính kế thừa lịch sử, tính quy luật phát sinh phát triển của hiện tượng, mối quan hệ biện chứng nhân quả giữa các hiện tượng cho nên điều kiện để xác lập mô hình dự báo là:

+ Các nguyên nhân, các yếu tố, các điều kiện cơ bản ảnh hưởng đến quy luật biến động phải tương đối ổn định, bền vững trong quá khứ đến hiện tại và tiến tới tương lai.

+ Một khi có sự thay đổi các yếu tố, các nguyên nhân thì phải xác định lại mô hình để thích nghi với hiện thực.

+ Để dễ điều chỉnh mô hình và đảm bảo mức độ chính xác phù hợp với thực tiễn thì tầm xa dự báo (là khoảng cách thời gian từ hiện tại đến tương lai) không nên quá $1/3$ thời gian của giai đoạn nghiên cứu dự báo.

- Tính khả thi của mức độ dự báo mang tính xác suất.

- Dự báo thống kê là dự báo ngắn hạn và dự báo trung hạn vì mức độ chính xác của kết quả dự báo thống kê tỷ lệ nghịch với tầm xa dự báo.

- Dự báo thống kê mang tính nhiều phương án. Cần phải lựa chọn phương án hay mô hình để làm hàm dự báo bằng cách kiểm định mô hình.

- Phương tiện để dự báo thống kê là các thuật toán, kỹ thuật tính toán phân tích, phương tiện tính toán, vi tính và trình độ nhận thức của người dự báo.

Để dự báo thống kê phải thực hiện tuần tự các bước sau:

- Phân tích thực trạng biến động của hiện tượng nghiên cứu bằng nhiều phương pháp thống kê để đánh giá bản chất, mối quan hệ nội tại của đối tượng nghiên cứu.

- Xác định mô hình dự báo, tính toán các tham số để định lượng chiều hướng, đáng điều biến động của tính quy luật.
- Kiểm định lựa chọn mô hình làm hàm dự báo.
- Phân tích hậu dự báo, theo dõi các yếu tố, nguyên nhân, điều kiện đã, đang và sẽ xảy ra, tham khảo ý kiến chuyên gia để điều chỉnh lại mô hình khi cần thiết.

6.4.2. Các phương pháp dự báo thống kê

6.4.2.1. Dự báo dựa vào lượng tăng (hoặc giảm) tuyệt đối trung bình

Phương pháp này được áp dụng trong trường hợp dãy số thời gian có các lượng tăng (hoặc giảm) tuyệt đối liên hoàn (từng kỳ) xấp xỉ nhau. Mức độ dự đoán ở thời kỳ thứ $n + L$ được xác định như sau:

$$\hat{y}_{n+L} = y_n + \overline{\Delta y} \cdot L$$

$$\overline{\Delta y} = \frac{y_n - y_1}{n - 1}$$

Trong đó: y_n - Là mức độ cuối cùng của dãy số;

L - Là tầm xa dự đoán;

$\overline{\Delta y}$ - Là lượng tăng (hoặc giảm) tuyệt đối trung bình.

6.4.2.2. Dự báo dựa vào tốc độ phát triển trung bình

Phương pháp này được sử dụng dùng trong trường hợp hiện tượng biến động với một nhịp độ tương đối ổn định, nghĩa là có tốc độ phát triển liên hoàn xấp xỉ nhau.

Mức độ dự đoán ở thời kỳ thứ $n + L$ là:

$$\hat{y}_{n+L} = y_n (\bar{t})^L$$

Trong đó: $\hat{y}_{n+L}$ - Mức độ dự đoán tại thời điểm $n + L$;

$\bar{t}$ - Tốc độ phát triển trung bình.

$$\bar{t} = \sqrt[n-1]{\frac{y_n}{y_1}}$$

y_n - Mức độ cuối cùng của dãy số;

y_1 - Mức độ đầu tiên của dãy số;

L - Tầm xa dự đoán.

6.4.2.3. Phương pháp ngoại suy hàm xu thế

Từ biến động thực tế của hiện tượng, xây dựng hàm hồi qui theo thời gian: $\bar{y}_t = f(t)$. Căn cứ vào hàm hồi qui đã xây dựng để dự đoán mức độ tương lai của hiện tượng.

Mức độ dự đoán ở thời điểm $(n + L)$ là:

$$\hat{y}_{n+L} = f(n+L)$$

Chương 7

CHỈ SỐ TRONG THỐNG KÊ

7.1. KHÁI NIỆM, Ý NGHĨA, PHÂN LOẠI CHỈ SỐ

7.1.1. Khái niệm chỉ số

Trong thống kê, chỉ số là một loại chỉ tiêu tương đối biểu hiện mối quan hệ so sánh giữa hai mức độ nào đó của một hiện tượng kinh tế. Như Doanh thu Bưu điện tỉnh A năm 2003 là 200 tỷ, năm 2004 là 240 tỷ đồng. Nếu lấy năm 2004 so với năm 2003 thì chỉ số phát triển doanh thu của Bưu điện Tỉnh A là 1,2 lần hay 120%.

Chỉ số được tính bằng cách so sánh hai mức độ của hiện tượng ở hai thời gian khác nhau nhằm nêu lên biến động của hiện tượng theo thời gian. Chỉ số này được biểu hiện bằng số tương đối giống như số tương đối động thái. Trong những mục đích so sánh khác, các chỉ số tính được cũng là số tương đối, có thể là số tương đối kế hoạch, số tương đối so sánh.

Như vậy trong thống kê khái niệm chỉ số tương đối rộng rãi; nó là phương pháp biểu hiện các quan hệ so sánh khác nhau. Trong thực tế, đối tượng chủ yếu của phương pháp phức tạp bao gồm nhiều đơn vị, nhiều phần tử có tính chất khác nhau về tên gọi, về giá trị sử dụng, đơn vị tính v.v... Muốn so sánh các mức độ của hiện tượng kinh tế phức tạp, cần biến đổi các đơn vị, các phần tử có tính chất khác nhau trở thành dạng đồng nhất.

7.1.2. Đặc điểm của chỉ số

Để đáp ứng yêu cầu so sánh các mức độ của hiện tượng phức tạp, chỉ số có đặc điểm là khi muốn so sánh các mức độ của hiện tượng kinh tế phức tạp, cần phải chuyển các đơn vị hoặc phần tử có tính chất khác nhau thành dạng giống nhau, để có thể cộng trực tiếp với nhau. Như khi các đơn vị có các sản phẩm dịch vụ khác nhau không thể trực tiếp cộng với nhau, do đó phải chuyển về dạng giống nhau (dạng giá trị). Như vậy dễ dàng so sánh hai chỉ tiêu giá trị để tính ra chỉ số giá trị.

Ngoài ra, chỉ số còn có đặc điểm là khi có nhiều nhân tố cùng tham gia vào việc tính toán số, phải giả định chỉ có một nhân tố thay đổi, còn các nhân tố khác không thay đổi.

7.1.3. Tác dụng của chỉ số

Chỉ số là một phương pháp không những có khả năng nêu lên biến động tổng hợp của hiện tượng phức tạp, mà còn có thể phân tích sự biến động này. Trong thống kê chỉ số có tác dụng:

- Chỉ số biểu hiện biến động của hiện tượng theo thời gian. Các chỉ số tính theo mục đích này được gọi là chỉ số phát triển. Nó được tính bằng cách so sánh mức độ của hiện tượng ở hai thời gian khác nhau.

- Biểu hiện sự biến động của hiện tượng theo không gian khác nhau giữa hai đơn vị, doanh nghiệp, hai địa phương. Chỉ số này gọi là chỉ số không gian.

- Biểu hiện nhiệm vụ và thực hiện kế hoạch của chỉ tiêu kinh tế nào đó. Chỉ số này gọi là chỉ số kế hoạch.

- Phân tích vai trò và ảnh hưởng của chỉ tiêu nhân tố đến sự biến động của chỉ tiêu tổng hợp.

7.1.4. Phân loại chỉ số

7.1.4.1. Căn cứ vào phạm vi tính toán

- *Chỉ số cá thể*: Biểu hiện sự biến động của từng phần tử, đơn vị cá biệt trong tổng thể phức tạp.

- *Chỉ số tổng hợp*: Biểu hiện sự biến động của tất cả các phần tử, các đơn vị của tổng thể nghiên cứu.

Cơ sở để tính chỉ số tổng hợp là chỉ số cá thể. Nó được dùng nhiều nhất trong thực tế.

7.1.4.2. Căn cứ vào tính chất của chỉ tiêu nghiên cứu

- *Chỉ số của chỉ tiêu số lượng*: Biểu hiện sự biến động của chỉ tiêu số lượng như khối lượng sản phẩm dịch vụ, số lượng công nhân, số lượng thiết bị.

- *Chỉ số của chỉ tiêu chất lượng*: Biểu hiện sự biến động của chỉ tiêu chất lượng như chỉ số giá thành sản phẩm, năng suất lao động v.v...

7.1.4.3. Căn cứ vào điều kiện thời gian và không gian

- Chỉ số thời gian như chỉ số phát triển, chỉ số kế hoạch, chỉ số định gốc, chỉ số liên hoàn.

- Chỉ số không gian biểu hiện quan hệ so sánh các mức độ của hiện tượng nghiên cứu giữa hai doanh nghiệp, hai địa phương...

Trong quá trình tính toán thường sử dụng các ký hiệu sau:

i - Chỉ số cá thể;

I - Chỉ số tổng hợp;

p - Biểu hiện mật lượng của chỉ tiêu chất lượng;

q - Biểu hiện mật lượng của chỉ tiêu số lượng.

Biểu hiện các mốc thời gian, sử dụng các ký hiệu:

0 - Kỳ gốc;

1 - Kỳ báo cáo hay kỳ thực hiện;

KH - Kỳ kế hoạch.

Biểu hiện các địa phương, các khu vực hoặc các đơn vị khác nhau, sử dụng các chữ cái A, B, C...

7.2. PHƯƠNG PHÁP TÍNH CHỈ SỐ

7.2.1. Chỉ số cá thể

7.2.1.1. Chỉ số cá thể của chỉ tiêu số lượng

Chỉ số cá thể của chỉ tiêu số lượng phản ánh sự biến động của từng phần tử, từng đơn vị có biểu hiện của hiện tượng nghiên cứu. Công thức tính:

$$i_q = \frac{q_1}{q_0}$$

Trong đó: q_1 - Mức độ của chỉ tiêu số lượng kỳ nghiên cứu;

q_0 - Mức độ của chỉ tiêu số lượng kỳ gốc.

Ví dụ: Giả sử có tài liệu thống kê về tình hình tiêu thụ sản phẩm của một doanh nghiệp Bưu điện như sau:

Bảng 7.1

Sản phẩm	Lượng tiêu thụ		Giá bán (1000đ)		Mức tiêu thụ (1000đ)		Chỉ số cá thể	
	Kỳ gốc q_0	Kỳ báo cáo q_1	Kỳ gốc p_0	Kỳ báo cáo p_1	Kỳ gốc p_0q_0	Kỳ báo cáo p_1q_1	i_q	i_p
A	1.000	1.100	500	450	500.000	495.000	1,1	0,9
B	2.000	2.400	300	240	600.000	576.000	1,2	0,8
C	4.000	6.000	400	400	1.600.000	2.400.000	1,5	1,0

Chỉ tiêu khối lượng là lượng sản phẩm tiêu thụ. Ta có các chỉ số cá thể:

$$i_q^A = \frac{q_1^A}{q_0^A} = \frac{1.100}{1.000} = 1,1 \text{ hay } 110 \%.$$

Tức là lượng sản phẩm A tiêu thụ kỳ báo cáo tăng 10% so với kỳ gốc.

$$i_q^B = \frac{q_1^B}{q_0^B} = \frac{2.400}{2.000} = 1,2 \text{ hay } 120 \%.$$

Tức là lượng sản phẩm B tiêu thụ kỳ báo cáo tăng 20% so với kỳ gốc.

$$i_q^C = \frac{q_1^C}{q_0^C} = \frac{6.000}{4.000} = 1,5 \text{ hay } 150 \%.$$

Tức là lượng sản phẩm C tiêu thụ kỳ báo cáo tăng 50% so với kỳ gốc.

7.2.1.2. Chỉ số cá thể của chỉ tiêu chất lượng

$$i_p = \frac{p_1}{p_0}$$

Trong đó: p_1, p_0 - Trị số của chỉ tiêu chất lượng kỳ báo cáo và kỳ gốc.

Với số liệu ở bảng, ta có thể tính chỉ số giá cả các sản phẩm:

- Đối với sản phẩm A:

$$i_p^A = \frac{p_1^A}{p_0^A} = \frac{450}{500} = 0,90 \text{ hay } 90\%.$$

Như vậy, giá bán sản phẩm A kỳ báo cáo giảm 10% so với kỳ gốc.

- Đối với sản phẩm B:

$$i_p^B = \frac{p_1^B}{p_0^B} = \frac{240}{300} = 0,80 \text{ hay } 80\%.$$

Như vậy, giá bán sản phẩm B kỳ báo cáo giảm 20% so với kỳ gốc.

- Đối với sản phẩm C:

$$i_p^C = \frac{p_1^C}{p_0^C} = \frac{400}{400} = 1,0$$

Như vậy, giá bán sản phẩm C không thay đổi kỳ báo cáo so với kỳ gốc.

7.2.2. Chỉ số tổng hợp

7.2.2.1. Nguyên tắc tính chỉ số tổng hợp

- Khi tính chỉ số tổng hợp phải chuyển biến các phân tử không đồng nhất không thể cộng lại với nhau của hiện tượng phức tạp về dạng đồng chất.

- Khi tính chỉ số tổng hợp của chỉ tiêu số lượng thì phải cố định chỉ tiêu chất lượng kỳ gốc.

- Khi tính chỉ số tổng hợp của chỉ tiêu chất lượng thì phải cố định chỉ tiêu số lượng kỳ báo cáo.

7.2.2.2. Phương pháp tính chỉ số tổng hợp của chỉ tiêu số lượng

Chỉ số tổng hợp của chỉ tiêu số lượng được tính theo công thức sau:

$$I_q = \frac{\sum_{i=1}^n p_{i(0)} q_{i(1)}}{\sum_{i=1}^n p_{i(0)} q_{i(0)}}$$

Ở đây chỉ tiêu chất lượng được chọn làm quyền số và cố định ở kỳ gốc.

Trở lại số liệu của bảng, ta có:

$$I_q = \frac{500 \times 1.100 + 300 \times 2.400 + 400 \times 6.000}{500.000 + 600.000 + 1.600.000} = 1,359$$

Như vậy, chỉ tiêu tiêu thụ chung của cả 03 sản phẩm kỳ báo cáo tăng 35,9% đã làm cho tổng mức tiêu thụ của cả 03 sản phẩm kỳ báo cáo tăng so với kỳ gốc một lượng là:

$$\begin{aligned} \Delta_{pq(q)} &= \sum p_{i(0)} q_{i(1)} - \sum p_{i(0)} q_{i(0)} = \\ &= 3.670.000 - 2.700.000 = 970.000 \text{ (ngàn đồng)} \end{aligned}$$

Chỉ số trung bình của chỉ tiêu số lượng:

Từ công thức tính chỉ số tổng hợp của chỉ tiêu số lượng qua phép biến đổi ta có:

$$I_q = \frac{\sum_{i=1}^n p_{i(0)} q_{i(1)}}{\sum_{i=1}^n p_{i(0)} q_{i(0)}} = \frac{\sum_{i=1}^n i_{qi} p_{i(0)} q_{i(0)}}{\sum_{i=1}^n p_{i(0)} q_{i(0)}}$$

Chỉ số tính theo công thức trên gọi là chỉ số trung bình số học của chỉ tiêu số lượng. Trọng số ở đây là $p_{i(0)} q_{i(1)}$.

7.2.2.3. Phương pháp tính chỉ số tổng hợp của chỉ tiêu chất lượng

Chỉ số tổng hợp của chỉ tiêu chất lượng được tính theo công thức:

$$I_p = \frac{\sum_{i=1}^n p_{i(1)} q_{i(1)}}{\sum_{i=1}^n p_{i(0)} q_{i(1)}}$$

Từ số liệu của bảng, ta tính được chỉ số tổng hợp giá cả của 03 sản phẩm:

$$\begin{aligned} I_p &= \frac{\sum p_{i(1)} q_{i(1)}}{\sum p_{i(0)} q_{i(1)}} = \frac{1.100 \times 450 + 2.400 \times 240 + 6.000 \times 400}{1.100 \times 500 + 2.400 \times 300 + 6.000 \times 400} \\ &= \frac{3.471.000}{3.670.000} = 0,945 \end{aligned}$$

Như vậy giá bán chung của cả 03 sản phẩm kỳ báo cáo giảm 5,5% so với kỳ gốc đã làm cho tổng mức tiêu thụ cả 03 sản phẩm kỳ báo cáo giảm so với kỳ gốc một lượng là:

$$\begin{aligned} \Delta_{pq(p)} &= \sum p_{i(1)} q_{i(1)} - \sum p_{i(0)} q_{i(1)} \\ &= 3.471.000 - 3.670.000 = - 201.000 \text{ (ngàn đồng)} \end{aligned}$$

Chỉ số trung bình của chỉ tiêu chất lượng:

Từ chỉ số tổng hợp của chỉ tiêu chất lượng bằng phép biến đổi ta có:

$$I_p = \frac{\sum_{i=1}^n p_{i(1)} q_{i(1)}}{\sum_{i=1}^n p_{i(0)} q_{i(1)}} = \frac{\sum_{i=1}^n p_{i(1)} q_{i(1)}}{\sum_{i=1}^n \frac{p_{i(1)}}{i_{pi}} q_{i(1)}}$$

Chỉ số tính theo công thức trên gọi là chỉ số trung bình của chỉ tiêu chất lượng.

7.2.3 Chỉ số không gian

7.2.3.1. Khái niệm chỉ số không gian

Là chỉ số biểu hiện quan hệ so sánh mức độ của một hiện tượng nghiên cứu trong điều kiện không gian khác nhau (các doanh nghiệp khác nhau hoặc trên các thị trường khác nhau).

7.2.3.2. Phương pháp tính chỉ số không gian

a) Tính chỉ số không gian của chỉ tiêu số lượng

Chỉ số không gian của chỉ tiêu số lượng biểu hiện mối quan hệ so sánh mức độ của hiện tượng nghiên cứu giữa hai doanh nghiệp hoặc hai địa phương khác nhau. Khi đó nếu q_A và q_B là giá trị của chỉ tiêu số lượng của hai doanh nghiệp A, B và p là giá trị của chỉ tiêu chất lượng có liên quan được chọn làm quyền số của chỉ số không gian của chỉ tiêu số lượng, ta có:

$$I_q(A/B) = \frac{\sum q_i^A p}{\sum q_i^B p}$$

Trong đó: p - Là chỉ tiêu chất lượng có liên quan và đóng vai trò là quyền số, vì vậy nó phải mang tính đặc trưng của cả hai doanh nghiệp. Có thể chọn p theo hai cách sau:

- Nếu giá trị của p đã được Nhà nước hoặc các Bộ, ngành qui định chung cho các doanh nghiệp thì lấy giá trị đó làm quyền số. Ví dụ như khi tính chỉ số khối lượng sản phẩm thì quyền số là giá cố định do Nhà nước qui định.

$$I_q(A/B) = \frac{\sum q_i^A p_{cd}}{\sum q_i^B p_{cd}}$$

Trong đó: p_{cd} - Là giá do Nhà nước ban hành.

- Nếu giá trị của p chưa thống nhất giữa các địa phương thì phải tính toán giá trị trung bình:

$$\bar{p}_i = \frac{p_i^A q_i^A + p_i^B q_i^B}{q_i^A + q_i^B}$$

$$I_q(A/B) = \frac{\sum q_i^A \bar{p}_i}{\sum q_i^B \bar{p}_i}$$

Trong đó: q_i^A, q_i^B - Là khối lượng hàng loại i bán ra trên thị trường A, B.

$\bar{p}_i$ - Là giá cả trung bình của mặt hàng i trên cả hai thị trường.

b) Chỉ số không gian của chỉ tiêu chất lượng

Khi xây dựng chỉ số không gian của chỉ tiêu chất lượng thì quyền số được chọn là chỉ tiêu số lượng có liên quan. Công thức tính:

$$I_p(A/B) = \frac{\sum p_i^A (q_i^A + q_i^B)}{\sum p_i^B (q_i^A + q_i^B)}$$

Trong đó: p_i^A, p_i^B - Là giá trị của chỉ tiêu chất lượng tại địa phương A, B. Ví dụ như giá cả loại hàng i trên thị trường A, B.

q_i^A, q_i^B - Là giá trị của chỉ tiêu số lượng có liên quan tại địa phương A, B. Ví dụ như khối lượng loại hàng i bán ra trên thị trường A, B.

Ta có mối liên hệ sau:

$$I_p(A/B) = \frac{1}{I_p(B/A)}$$

Ví dụ: Có tài liệu về khối lượng sản phẩm dịch vụ tiêu thụ và giá bán các sản phẩm của hai đơn vị (trên hai địa phương khác nhau) như sau. Tính chỉ số không gian về giá cả và lượng sản phẩm dịch vụ tiêu thụ hai đơn vị A và B

Bảng 7.2

Sản phẩm dịch vụ	Đơn vị A		Đơn vị B		$q_i^A + q_i^B$
	q_i^A	p_i^A	q_i^B	p_i^B	
X	450	30	550	32	1.000
Y	320	28	280	25	600
Z	840	22	660	20	1.500

Ta có:

$$\begin{aligned}
 I_p(A/B) &= \frac{\sum p_i^A (q_i^A + q_i^B)}{\sum p_i^B (q_i^A + q_i^B)} = \\
 &= \frac{30 \times 1.000 + 28 \times 600 + 22 \times 1.500}{32 \times 1.000 + 25 \times 600 + 20 \times 1.500} = 1,036
 \end{aligned}$$

Nghĩa là giá bán ở đơn vị A cao hơn đơn vị B là 3,6%.

$$I_q(A/B) = \frac{\sum q_i^A \bar{p}_i}{\sum q_i^B \bar{p}_i}$$

Giá bán trung bình của các sản phẩm dịch vụ là:

$$\bar{p}_x = \frac{30 \times 450 + 32 \times 550}{450 + 550} = 31,10$$

$$\bar{p}_y = \frac{28 \times 320 + 25 \times 280}{320 + 280} = 26,60$$

$$\bar{p}_z = \frac{22 \times 840 + 20 \times 660}{840 + 660} = 21,12$$

Do đó:

$$I_q(A/B) = \frac{450 \times 31,1 + 26,6 \times 320 + 21,12 \times 840}{550 \times 31,1 + 280 \times 26,6 + 660 \times 21,12} = 1,045$$

Có nghĩa là lượng sản phẩm dịch vụ tiêu thụ tại đơn vị A nhiều hơn đơn vị B là 4,5%.

7.2.4. Phương pháp tính chỉ số của chỉ tiêu bình quân

Chỉ số của chỉ tiêu bình quân được xác định như sau:

$$\bar{I}_x = \frac{\bar{X}_1}{\bar{X}_0} = \frac{\frac{\sum X_1 f_1}{f_0}}{\frac{\sum X_0 f_0}{f_0}}$$

Trong đó: X_1, X_0 - Lượng biến của tiêu thức kỳ báo cáo và kỳ gốc;

f_1, f_0 - Số đơn vị tổng thể kỳ báo cáo và kỳ gốc.

$\bar{X}_1, \bar{X}_0$ - Số bình quân số học của các lượng biến kỳ báo cáo và kỳ gốc.

Chỉ số tính theo công thức trên gọi là chỉ số phát triển của chỉ tiêu bình quân.

Khi lượng biến thay đổi, ta có chỉ số sau:

$$\bar{I}_x = \frac{\frac{\sum X_1 f_1}{\sum f_1}}{\frac{\sum X_0 f_1}{\sum f_1}}$$

Chỉ số tính theo công thức này được gọi là chỉ số cấu thành khả biến.

Khi kết cấu (số đơn vị) của tổng thể thay đổi, ta có chỉ số sau:

$$\bar{I}_k = \frac{\frac{\sum X_0 f_1}{\sum f_1}}{\frac{\sum X_0 f_0}{\sum f_0}}$$

Chỉ số tính theo công thức này gọi là chỉ số ảnh hưởng kết cấu.

Ta có mối quan hệ sau:

Chỉ số phát triển = Chỉ số cấu thành khả biến x Chỉ số ảnh hưởng kết cấu

$$\frac{\frac{\sum X_1 f_1}{\sum f_1}}{\frac{\sum X_0 f_0}{\sum f_0}} = \frac{\frac{\sum X_1 f_1}{\sum f_1}}{\frac{\sum X_0 f_1}{\sum f_1}} = \frac{\frac{\sum X_0 f_1}{\sum f_1}}{\frac{\sum X_0 f_0}{\sum f_0}}$$

7.3. HỆ THỐNG CHỈ SỐ

7.3.1. Hệ thống chỉ số

Như ta đã trình bày ở trên là phương pháp chỉ số giúp nghiên cứu sự thay đổi của hiện tượng kinh tế - xã hội qua thời gian và

không gian. Ngoài ra phương pháp chỉ số còn có thể phân tích mối liên hệ giữa các nhân tố và mức độ ảnh hưởng của các nhân tố đến sự thay đổi của một chỉ tiêu kinh tế tổng hợp. Ví dụ mức tiêu thụ hàng hóa của một công ty biến động là do ảnh hưởng của hai nhân tố giá bán và lượng hàng tiêu thụ. Tổng chi phí sản xuất của các sản phẩm sản xuất ra là do ảnh hưởng của giá thành và khối lượng sản phẩm sản xuất, khối lượng sản phẩm sản xuất ra phụ thuộc vào năng suất lao động và số công nhân...

Giả sử ta cần phân tích ảnh hưởng của tổng mức tiêu thụ hàng hóa qua hai thời kỳ nghiên cứu trong mối liên hệ với hai nhân tố: giá cả và khối lượng hàng hóa tiêu thụ. Hệ thống chỉ số phản ánh mối liên hệ giữa các chỉ tiêu trên như sau:

$$I_{pq} = I_p \times I_q$$

Trong đó: I_{pq} - Chỉ số tổng hợp mức hàng hóa tiêu thụ;

I_p - Chỉ số tổng hợp giá cả;

I_q - Chỉ số tổng hợp khối lượng.

Hay:

$$\frac{\sum_1^n p_{i(1)} q_{i(1)}}{\sum_1^n p_{i(0)} q_{i(0)}} 100 = \frac{\sum_1^n p_{i(1)} q_{i(1)}}{\sum_1^n p_{i(0)} q_{i(1)}} 100 \times \frac{\sum_1^n p_{i(0)} q_{i(1)}}{\sum_1^n p_{i(0)} q_{i(0)}} 100$$

Hệ thống chỉ số $I_{pq} = I_p \times I_q$ có thể được vận dụng để phân tích cho các chỉ tiêu kinh tế tổng hợp khác do ảnh hưởng của các nhân tố có liên quan. Trong đó:

I_{pq} - Chỉ số tổng hợp phản ánh sự thay đổi của một chỉ tiêu kinh tế nào đó do ảnh hưởng đồng thời của cả hai nhân tố p và q . Trong đó cần lưu ý tới ý nghĩa kinh tế của tích số $p \times q$.

I_p - Chỉ số chỉ tiêu chất lượng (như giá cả, giá thành sản phẩm, năng suất lao động,...).

I_q - Chỉ số chỉ tiêu khối lượng (như khối lượng hàng hóa tiêu thụ, khối lượng sản phẩm sản xuất, số công nhân...).

7.3.2. Vận dụng phương pháp chỉ số phân tích ảnh hưởng của các nhân tố đến chỉ tiêu kinh tế tổng hợp

Để phân tích ảnh hưởng này, người ta có thể sử dụng hệ thống chỉ số.

Ví dụ: Có tài liệu về lượng sản phẩm dịch vụ tiêu thụ và giá cả các loại sản phẩm dịch vụ như sau:

Bảng 7.3

Loại sản phẩm dịch vụ	Giá cả (ngàn đồng)		Lượng tiêu thụ	
	Kỳ gốc	Kỳ báo cáo	Kỳ gốc	Kỳ báo cáo
A	3	5	50	190
B	1,6	2,4	100	120
C	2,4	3,6	200	360
D	40	25	4	4,2

Yêu cầu phân tích ảnh hưởng của lượng tiêu thụ sản phẩm dịch vụ và giá cả đến tổng mức tiêu thụ sản phẩm dịch vụ.

Ta có hệ thống chỉ số sau:

$$\frac{\sum_1^n p_{i(1)} q_{i(1)}}{\sum_1^n p_{i(0)} q_{i(0)}} 100 = \frac{\sum_1^n p_{i(1)} q_{i(1)}}{\sum_1^n p_{i(0)} q_{i(1)}} 100 \times \frac{\sum_1^n p_{i(0)} q_{i(1)}}{\sum_1^n p_{i(0)} q_{i(0)}} 100$$

Để phân tích ảnh hưởng của các nhân tố giá cả và lượng sản phẩm dịch vụ tiêu thụ đến tổng mức tiêu thụ sản phẩm, ta lập bảng sau:

Bảng 7.4

Loại sản phẩm	Giá cả		Lượng hàng tiêu thụ		p_0q_0	p_1q_1	p_0q_1
	p_0	p_1	q_0	q_1			
A	3	5	50	190	150	950	570
B	1,6	2,4	100	120	160	288	192
C	2,4	3,6	200	360	480	1.296	864
D	40	25	4	4,2	160	105	168
Cộng					950	2.639	1.794

Căn cứ vào số liệu đã tính trong bảng, ta có:

$$\frac{2.639}{950}100 = \frac{2.639}{1.794}100 \times \frac{1.794}{950}100$$

Hay:

$$277,8\% = 147,1\% \times 188,8\%$$

Và:

$$\begin{aligned} & \left(\sum_1^n p_{i(1)} q_{i(1)} - \sum_1^n p_{i(0)} q_{i(0)} \right) = \\ & = \left(\sum_1^n p_{i(1)} q_{i(1)} - \sum_1^n p_{i(0)} q_{i(1)} \right) + \left(\sum_1^n p_{i(0)} q_{i(1)} - \sum_1^n p_{i(0)} q_{i(0)} \right) \\ & (2.639 - 950) = (2.639 - 1.794) + (1.794 - 950) \\ & 1.689 = 845 + 844 \text{ (ngàn đồng)} \end{aligned}$$

Phân tích ảnh hưởng của các nhân tố tính theo tỷ lệ phần trăm:

$$\begin{aligned} & \frac{\sum_1^n p_{i(1)} q_{i(1)} - \sum_1^n p_{i(0)} q_{i(0)}}{\sum_1^n p_{i(0)} q_{i(0)}} 100 = \\ &= \frac{\sum_1^n p_{i(1)} q_{i(1)} - \sum_1^n p_{i(0)} q_{i(1)}}{\sum_1^n p_{i(0)} q_{i(1)}} 100 \times \frac{\sum_1^n p_{i(0)} q_{i(1)} - \sum_1^n p_{i(0)} q_{i(0)}}{\sum_1^n p_{i(0)} q_{i(0)}} 100 \\ & \frac{1.689}{950} 100 = \frac{845}{850} 100 \times \frac{844}{950} 100 \end{aligned}$$

Hay: $177,8\% = 90\% + 88,8\%$.

Kết quả tính toán trên cho phép ta kết luận: Tổng mức tiêu thụ sản phẩm dịch vụ tính chung cho cả bốn loại kỳ báo cáo so với kỳ gốc tăng 177,8%, với mức tăng tuyệt đối là 1.689 ngàn đồng là do ảnh hưởng của hai nhân tố có liên quan:

- Thứ nhất: Do giá cả kỳ báo cáo so với kỳ gốc tăng 47,1% làm cho tổng mức tiêu thụ sản phẩm dịch vụ tăng 90%, mức tăng tuyệt đối là 845 ngàn đồng.

- Thứ hai: Do khối lượng sản phẩm dịch vụ tiêu thụ kỳ báo cáo so với kỳ gốc tăng 88,8%, làm cho tổng mức tiêu thụ sản phẩm tăng 88,8%, ứng với một lượng tăng tuyệt đối là 844 ngàn đồng.

Phụ lục
CÁC BẢNG SỐ

Bảng 1

Giá trị của hệ số $C_m^x = \frac{m!}{x!(m-x)!}$

$\begin{matrix} x \\ m \end{matrix}$	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	1	2	1													
3	1	3	3	1												
4	1	4	6	4	1											
5	1	5	10	10	5	1										
6	1	6	15	20	15	6	1									
7	1	7	21	35	35	21	7	1								
8	1	8	28	56	70	56	28	8	1							
9	1	9	36	84	126	126	84	36	9	1						
10	1	10	45	120	210	252	210	120	45	10	1					
11	1	11	55	165	330	462	462	330	165	55	11	1				
12	1	12	66	220	495	792	924	792	495	220	66	12	1			
13	1	13	78	286	705	1287	1716	1716	1287	715	286	78	13	1		
14	1	14	91	364	1001	2002	3003	3432	3003	2002	1001	364	91	14	1	
15	1	15	105	455	1365	3003	5005	4435	6435	5005	3003	364	45	105	15	1

Bảng 2 (tiếp theo)

$\lambda \backslash x$	3,1	3,2	3,3	3,4	3,5	3,6	3,7	3,8	3,9	4,0
0	0,0450	0408	0369	0334	0302	0273	0247	0224	0202	0146
1	1397	1304	1217	1135	1057	0984	0915	0850	0789	0733
2	2165	2087	2008	1929	1850	1771	1692	1615	1539	1465
3	2237	2226	2209	2186	2158	2125	2087	2046	2001	1954
4	1734	1781	1823	1858	1888	1932	1931	1944	1951	1954
5	1075	1140	1203	1264	1322	1377	1429	1477	1522	1563
6	0555	0608	0662	0716	0771	0826	0881	0936	0989	1042
7	0246	0278	0312	0348	0385	0425	0466	0508	0551	0595
8	0095	0111	0129	0148	0109	0295	0215	0241	0269	0298
9	0033	0040	0044	0056	0066	0076	0089	0102	0116	0132
10	0010	0013	0016	0019	0023	0028	0033	0039	0045	0053
11	0008	0004	0005	0006	0007	0009	0011	0013	0016	0019
12	0001	0001	0001	0002	0002	0003	0003	0004	0005	0006
13	0000	0000	0000	0000	0001	0001	0001	0001	0002	0002
14	0000	0000	0000	0000	0000	0000	0000	0000	0000	0001
$\lambda \backslash x$	4,1	4,2	4,3	4,4	4,5	4,6	4,7	4,8	4,9	5,0
0	0,0166	0150	0136	0123	0111	0101	0091	0082	0074	0067
1	0679	0630	0583	0540	0500	0462	0427	0395	0365	0337
2	1393	1323	1254	1188	1125	1063	1005	0948	0894	0842
3	1904	1852	1798	1743	1687	1631	1574	1517	1460	1404
4	1951	1944	1933	1917	1898	1875	1849	1820	1789	1755
5	1600	1633	1662	1687	1708	1725	1738	1747	1753	1753
6	1093	1143	1191	1237	1281	1323	1362	1398	1432	1460
7	0640	0606	0732	0778	0824	0869	0914	0959	1002	1044
8	0388	0360	0393	0428	0463	0500	0537	0575	0614	0653
9	0150	0168	0188	0209	0232	0250	0280	0307	0334	0365
10	0061	0071	0081	0092	0104	1008	0132	0147	0164	0187
11	0023	0027	0032	0037	0043	0049	0056	0064	0073	0083
12	0008	0009	0011	0014	0016	0019	0022	0026	0030	0034
13	0002	0003	0004	0005	0006	0007	0008	0009	0011	0013
14	0001	0001	0001	0001	0002	0002	0003	0003	0004	0005
15	0000	0000	0000	0000	0001	0001	0001	0001	0001	0002

Bảng 2 (tiếp theo)

$\lambda \backslash x$	5,1	5,2	5,3	5,4	5,5	5,6	5,7	5,8	5,9	6,0
0	0,0061	0055	0030	0015	0041	0037	0033	0030	0027	0025
1	0311	0287	0265	0244	0025	0207	0191	0176	0162	0149
2	0793	0746	0701	0659	0618	0580	0544	0509	0477	0446
3	1348	1293	1239	1185	1138	1082	1033	0985	0938	0992
4	1719	1681	1641	1600	1558	1515	1472	1428	1333	1339
5	1753	1748	1740	1728	1714	1697	1678	1656	1632	1606
6	1490	1515	1537	1555	1571	1584	1594	1601	1605	1606
7	1086	1125	1163	1200	1234	1267	1290	1386	1353	1377
8	0692	0731	0771	0810	0849	0087	0925	0962	0998	1033
9	0392	0423	0454	0486	0519	0551	0586	0620	0654	0688
10	0200	0220	0241	0262	0285	0309	0334	0359	0386	0413
11	0093	0104	0116	0129	0143	0157	0173	0190	0207	0225
12	0039	0045	0051	0058	0065	0073	0082	0092	0102	0113
13	0015	0018	0021	0024	0028	0032	0036	0041	0046	0052
14	0006	0007	0008	0009	0011	0013	0015	0017	0019	0022
15	0002	0002	0003	0003	0004	0005	0006	0007	0008	0009
16	0002	0001	0001	0001	0001	0002	0002	0003	0003	0003
17	0000	0000	0000	0000	0000	0001	0001	0001	0001	0001
$\lambda \backslash x$	6,1	6,2	6,3	6,4	6,5	6,6	6,7	6,8	6,9	7,0
0	0,0022	0020	0018	0017	0015	0014	0012	0011	0010	0009
1	0137	0126	0116	0106	0098	0090	0082	0076	0071	0061
2	0417	0390	0364	0340	0318	0296	0276	0258	0070	0064
3	0848	0806	0765	0726	0688	0552	0617	0594	0552	0521
4	1294	1249	1205	1162	1118	1076	1034	0992	0952	0912
5	1579	1549	1519	1487	1454	1420	1385	1349	1314	1277
6	1605	1601	1595	1586	1575	1562	1546	1529	1511	1490
7	1399	1418	1405	1450	1462	1472	1480	1486	1489	1490
8	1066	1099	1130	1160	1188	1215	1240	1263	1284	1304
9	0723	0751	0791	0825	0585	0891	0923	0954	0985	1014
10	0441	0469	0498	0528	0558	0588	0618	0649	0679	0710
11	0245	0265	0285	0307	0330	0353	0377	0401	0426	0452
12	0124	0137	0150	0164	0179	0194	0210	0227	0245	0264
13	0058	0065	0073	0081	0089	0098	0108	0119	0130	0142
14	0025	0029	0033	0037	0041	0046	0052	0058	0064	0071
15	0010	0012	0014	0016	0018	0020	0023	0026	0029	0033
16	0004	0005	0005	0006	0007	0008	0010	0011	0013	0014
17	0001	0002	0002	0002	0003	0003	0004	0004	0005	0006
18	0000	0001	0002	0001	0001	0001	0001	0002	0002	0002
19	0000	0000	0000	0000	0000	0000	0000	0001	0001	0001

Bảng 2 (tiếp theo)

$\lambda \backslash x$	7,1	7,2	7,3	7,4	7,5	7,6	7,7	7,8	7,9	8,0
0	0,0080	0007	0007	0006	0006	0005	0005	0004	0004	0003
1	0059	0054	0049	0045	0041	0038	0035	0032	0029	0027
2	0208	0194	0180	0167	0156	0145	0134	0125	0116	0107
3	0492	0464	0438	0413	0389	0366	0345	0324	0305	0286
4	0074	0836	0799	0764	0729	0696	0663	0632	0602	0573
5	1241	1204	1167	1130	01094	1057	1021	0986	0951	0916
6	1468	1445	1420	1394	1367	1339	1391	1290	1252	1001
7	1489	1486	1481	1474	1465	1454	1442	1400	1413	1396
8	1321	1337	1351	1363	1373	1382	1388	1391	1395	1396
9	1042	1070	1096	1121	1144	1167	1187	1207	1224	1241
10	0740	0770	0800	0829	0858	0887	0914	0941	0967	0993
11	0478	0504	0531	0558	0585	0613	0640	0667	0695	0722
12	0283	0303	0323	0344	0366	0388	0411	0434	0457	0481
13	0054	0168	0181	0196	0211	0227	0243	0260	0278	0296
14	0078	0086	0095	0104	0113	0123	0134	0145	0157	0169
15	0037	0041	0046	0051	0057	0062	0069	0075	0083	0090
16	0016	0019	0021	0024	0026	0030	0033	0037	0041	0045
17	0007	0008	0009	0010	0012	0013	0015	0017	0019	0021
18	0003	0003	0004	0004	0005	0006	0006	0007	0008	0009
19	0001	0001	0001	0002	0002	0002	0003	0003	0003	0004
20	0000	0000	0001	0001	0001	0001	0001	0001	0001	0002
21	0000	0000	0000	0000	0000	0000	0000	0000	0001	0001
$\lambda \backslash x$	8,1	8,2	8,3	8,4	8,5	8,6	8,7	8,8	8,9	9,0
0	0,003	0003	0002	0002	0002	0002	0002	0002	0001	0001
1	0025	0023	0021	0019	0017	0016	0014	0013	0012	0011
2	0100	0092	0086	0079	0074	0068	0063	0058	0054	0050
3	0269	0252	0237	0222	0208	0195	0183	0171	0160	0150
4	0544	0517	0491	0466	0443	0420	0398	0377	0375	0337
5	0882	0849	0816	0784	0752	0722	0692	0663	0635	0607
6	1191	1160	1128	1097	1066	1034	1003	0972	0941	0911
7	1378	1358	1338	1317	1294	1271	1247	1222	1197	1171
8	1395	1392	1388	1382	1375	1366	1356	1344	1332	1318
9	1256	1269	1280	1290	1299	1306	1311	1315	1317	1318
10	1017	1040	1063	1084	1104	1123	1140	1157	1172	1186
11	0749	0776	0802	0828	0853	0878	0902	0925	0948	0970
12	0505	0530	0555	0579	0604	0629	0654	0679	0703	0728
13	0315	0334	0354	0374	0395	0416	0438	0459	0481	0504
14	0182	0196	0210	0225	0240	0256	0272	0289	0306	0324
15	0098	0107	0116	0126	0136	0147	0158	0169	0182	0194
16	0050	0055	0060	0066	0072	0079	0080	0093	0101	0109
17	0024	0026	0029	0033	0036	0040	0044	0048	0053	0058
18	0011	0012	0014	0015	0017	0019	0021	0024	0026	0029
19	0005	0005	0006	0007	0008	0009	0010	0011	0012	0014
20	0002	0002	0002	0003	0003	0004	0004	0005	0005	0006
21	0001	0001	0001	0001	0001	0002	0002	0002	0002	0003
22	0000	0000	0000	0000	0001	0001	0001	0001	0001	0001

Bảng 2 (tiếp theo)

λ x	9,1	9,2	9,3	9,4	9,5	9,6	9,7	9,8	9,9	10,0
0	0,0001	0001	0001	0001	0001	0001	0001	0001	0001	0000
1	0010	0009	0009	0008	0007	0007	0006	0005	0005	0005
2	0046	0043	0040	0037	0034	0031	0029	0027	0025	0023
3	0140	0131	0123	0115	0107	0100	0093	0087	0081	0076
4	0319	0302	0285	0269	0254	0240	0226	0213	0201	0189
5	0581	0555	0530	0506	0483	0460	0439	0418	0390	0378
6	0881	0851	0822	0793	0764	0736	0709	0682	0656	0631
7	1145	1118	1091	1064	1037	1060	0982	0955	0928	0631
8	1302	1286	1268	1251	1232	1212	1191	1170	1148	1126
9	1317	1315	1306	1300	1293	1281	1284	1274	1263	1251
10	1198	1210	1249	1228	1235	1241	1245	1249	1250	1251
11	0991	1012	1031	1049	1067	1083	1098	1112	1125	1137
12	0752	0776	0799	0822	0844	0866	0888	0908	0928	0948
13	0526	0549	0572	0594	0617	0640	0662	0685	0707	0729
14	0342	0361	0380	0399	0419	0439	0459	0479	0500	0521
15	0208	0221	0235	0250	0265	0281	0297	0313	0330	0347
16	0118	0127	0137	0147	0157	0168	0180	0192	0204	0217
17	0063	0069	0075	0081	0088	0095	0103	0111	0119	0128
18	0032	0035	0039	0042	0046	0051	0055	0060	0065	0071
19	0015	0017	0019	0021	0023	0026	0028	0031	0034	0037
20	0007	0008	0009	0010	0011	0012	0014	0015	0017	0019
21	0003	0003	0004	0004	0005	0006	0006	0007	0008	0009
22	0001	0001	0002	0002	0002	0002	0003	0003	0004	0004
23	0000	0001	0001	0001	0001	0001	0001	0001	0002	0002
24	0000	0000	0000	0000	0000	0000	0000	0001	0001	0001

Bảng 3: Bảng giá trị của χ^2

χ^2 / P	0,05	0,02	0,01	0,001
1	3,84	5,41	6,64	10,83
2	5,99	7,82	9,21	13,82
3	7,82	9,84	11,34	16,22
4	9,49	11,67	13,28	18,46
5	11,07	13,39	15,09	20,52
6	12,59	15,03	16,81	22,46
7	14,07	16,62	18,48	24,32
8	15,51	18,17	20,09	26,12
9	16,92	19,68	21,67	27,88
10	18,31	21,16	23,21	29,59
11	19,68	22,62	24,72	31,26
12	21,03	24,05	26,22	32,91
13	22,36	25,47	27,69	33,91
14	23,68	26,87	29,14	34,53
15	25,00	28,26	30,58	37,70
16	26,30	29,63	32,00	39,25
17	27,59	30,63	33,00	40,25
18	28,87	32,35	34,80	42,31
19	30,14	33,69	36,19	43,82
20	31,41	35,02	37,57	45,32
21	32,67	36,34	38,93	46,80
22	33,92	37,66	40,29	48,27
23	35,17	38,97	41,64	49,73
24	36,42	40,27	42,98	51,18
25	37,67	41,57	44,31	52,62
26	38,88	42,86	45,64	54,05
27	40,41	44,14	46,96	55,48
28	41,34	45,42	48,28	56,89
29	42,56	46,69	49,59	58,30
30	43,77	47,96	50,89	59,70

Bảng 4: Bảng giá trị của D

$n \backslash p$	0,05	0,01	$n \backslash p$	0,05	0,01
1	0,975	0,995	21	0,287	0,344
2	0,842	0,929	22	0,281	0,337
3	0,708	0,829	23	0,275	0,330
4	0,624	0,734	24	0,269	0,323
5	0,563	0,669	25	0,264	0,317
6	0,519	0,617	26	0,259	0,311
7	0,483	0,576	27	0,254	0,305
8	0,454	0,542	28	0,250	0,300
9	0,430	0,513	29	0,246	0,295
10	0,409	0,489	30	0,242	0,290
11	0,391	0,468	31	0,238	0,285
12	0,375	0,441	32	0,234	0,281
13	0,361	0,432	33	0,231	0,277
14	0,349	0,418	34	0,227	0,275
15	0,338	0,404	35	0,224	0,269
16	0,327	0,392	36	0,221	0,265
17	0,318	0,381	37	0,218	0,262
18	0,309	0,371	38	0,215	0,258
19	0,301	0,361	39	0,213	0,255
20	0,294	0,352	40	0,210	0,252
			> 40	$1,36/\sqrt{4}$	$1,63/\sqrt{4}$

Bảng 5

$$\text{Giá trị của hàm } \varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

t	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3965	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3449	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2510	2492	2468	2444
1,0	2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1519
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1113
1,6	1100	1092	1174	1057	1040	1023	1000	0989	0873	0852
1,7	0940	0925	0909	0893	0800	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0388	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0174	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0078	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0048	0046
3,0	0044	0043	0042	0040	0039	0038	0037	0036	0035	0034

Bảng 6: Bảng hệ số tương quan r

P Độ tự do	0,10	0,05	0,02	0,01
1	0,9877	0,9969	0,9995	0,9999
2	0,9000	0,9500	0,9800	0,9900
3	0,8054	0,8783	0,9343	0,9587
4	0,7293	0,8114	0,8822	0,9172
5	0,6694	0,7545	0,8329	0,8745
6	0,6215	0,7067	0,7887	0,8343
7	0,5822	0,6640	0,7498	0,7977
8	0,5491	0,6319	0,7155	0,7646
9	0,5214	0,6021	0,6851	0,7348
10	0,4973	0,5760	0,6581	0,7079
11	0,4762	0,5529	0,6339	0,6835
12	0,4575	0,5324	0,6120	0,6614
13	0,4409	0,5239	0,5923	0,6411
14	0,4259	0,4973	0,5742	0,6226
15	0,4124	0,4821	0,5577	0,6055
16	0,4000	0,4683	0,5425	0,5897
17	0,3887	0,4555	0,5285	0,5751
18	0,3783	0,4438	0,5155	0,5614
19	0,3687	0,4329	0,5034	0,5487
20	0,3598	0,4227	0,4921	0,5368
25	0,3233	0,3809	0,4451	0,4869
30	0,2960	0,3494	0,4093	0,4487
35	0,2746	0,3246	0,3810	0,4182
40	0,2573	0,3044	0,3578	0,3932
45	0,2428	0,2875	0,3384	0,3721
50	0,2306	0,2732	0,3218	0,3541
60	0,2108	0,2500	0,2948	0,3248
70	0,1954	0,2319	0,2737	0,3017
80	0,1829	0,2172	0,2565	0,2830
90	0,1726	0,2050	0,2422	0,2673
100	0,1638	0,1946	0,2301	0,2540

TÀI LIỆU THAM KHẢO

1. PGS.TS Bùi Xuân Phong - *Thống kê vận tải* - Trường Đại học Giao thông Vận tải, 1999.
2. PGS.TS Bùi Xuân Phong - *Thống kê và ứng dụng* - Nhà xuất bản Thống kê, 2002.
3. Hà Văn Sơn - *Giáo trình Lý thuyết thống kê ứng dụng trong quản trị và kinh tế* - Nhà xuất bản Thống kê, 2004.
4. PGS.TS Tô Phi Phương - *Giáo trình Lý thuyết thống kê* - Nhà xuất bản Giáo dục, 1998.
5. TS. Trần Văn Thắng - *Giáo trình lý thuyết thống kê* - Nhà xuất bản Thống kê, 1998.
6. TS. Nguyễn Huy Thịnh - *Lý thuyết thống kê* - Nhà xuất bản Tài chính, 1998.

MỤC LỤC

<i>Lời Nhà xuất bản</i>	5
Chương 1: MỘT SỐ KHÁI NIỆM VỀ THỐNG KÊ	
1.1. Thống kê học là gì?	7
1.2. Tổng thể thống kê và đơn vị tổng thể thống kê	8
1.3. Tiêu thức thống kê	10
1.4. Chỉ tiêu thống kê	13
1.5. Các loại thang đo.....	14
Chương 2: CÁC PHƯƠNG PHÁP TRÌNH BÀY SỐ LIỆU THỐNG KÊ	
2.1. Trình bày số liệu bằng bảng thống kê	16
2.2. Trình bày số liệu bằng đồ thị	24
Chương 3: CÁC MỨC ĐỘ CỦA HIỆN TƯỢNG THỐNG KÊ	
3.1. Số tuyệt đối	28
3.2. Số tương đối	31
3.3. Số bình quân	36
3.4. Các chỉ tiêu đánh giá độ biến thiên của tiêu thức	51
3.5. Phân phối trong thống kê	62
Chương 4: ĐIỀU TRA CHỌN MẪU	
4.1. Khái niệm, ưu nhược điểm và phạm vi sử dụng điều tra chọn mẫu	79
4.2. Điều tra chọn mẫu ngẫu nhiên	81
4.3. Điều tra chọn mẫu phi ngẫu nhiên	104

Chương 5: TƯƠNG QUAN VÀ HỒI QUY

5.1. Khái niệm.....	107
5.2. Liên hệ tương quan tuyến tính giữa hai tiêu thức số lượng	109
5.3. Liên hệ tương quan phi tuyến giữa hai tiêu thức số lượng	114
5.4. Liên hệ tương quan tuyến tính giữa nhiều tiêu thức số lượng	118

Chương 6: DÃY SỐ THỜI GIAN

6.1. Khái niệm, phân loại và ý nghĩa của dãy số thời gian.....	121
6.2. Các chỉ tiêu phân tích dãy số thời gian	123
6.3. Các phương pháp biểu hiện xu hướng phát triển của hiện tượng	130
6.4. Dự báo thống kê ngắn hạn	140

Chương 7: CHỈ SỐ TRONG THỐNG KÊ

7.1. Khái niệm, ý nghĩa, phân loại chỉ số.....	145
7.2. Phương pháp tính chỉ số.....	148
7.3. Hệ thống chỉ số	157

<i>Phụ lục</i>	162
----------------------	-----

<i>Tài liệu tham khảo</i>	173
---------------------------------	-----

THỐNG KÊ VÀ ỨNG DỤNG TRONG BƯU CHÍNH, VIỄN THÔNG

Chịu trách nhiệm xuất bản
LƯU ĐỨC VĂN

Biên tập: TRẦN CHÍ ĐẠT
NGÔ TẤN ĐẠT
Chế bản: NGUYỄN MẠNH HOÀNG
Sửa bản in: NGÔ TẤN ĐẠT
Trình bày bìa: NGUYỄN MẠNH HOÀNG

NHÀ XUẤT BẢN BƯU ĐIỆN

Trụ sở: 18 Nguyễn Du, TP. Hà Nội
Điện thoại: 04-9430202, 04-9432438 **Fax:** 04-9431285
E-mail: bientap@hn.vnn.vn
Chi nhánh: 27 Nguyễn Bỉnh Khiêm, Quận I, TP. Hồ Chí Minh
Điện thoại: 08-9100925 **Fax:** 08-9100924
E-mail: chinhanh-nxbdd@hcm.vnn.vn